Population structure and connectivity of shellfisheries resources in the Irish Sea

Hayley Victoria Watson

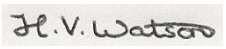
A thesis submitted in candidature for the degree of

Doctor of Philosophy

September 2014

DECLARATION

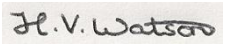
This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed  (Candidate)

Date .....12.06.2015.....

STATEMENT 1

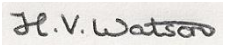
This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by giving explicit references and a full bibliography is appended.

Signed  (Candidate)

Date .....12.06.2015.....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed  (Candidate)

Date .....12.06.2015.....

ABSTRACT

The stock structure of edible crab (*Cancer pagurus*) and European lobster (*Homarus gammarus*) in the Irish Sea was investigated using eleven and twelve microsatellite markers, respectively. These shellfish species are of high economic value to the UK and Ireland and are being exploited with increasing intensity due to the decline of fin-fisheries. Population structure of *H. gammarus* was assessed by combining population genetic analysis of adults and biophysical modelling of larval dispersal, with particular emphasis on assessing recruitment patterns for the Lundy (South Wales) No-Take Zone (NTZ). Genetic structuring was statistically non-significant, compatible with recurrent spatial connectivity predicted by the larval dispersal modelling. The NTZ exhibited a statistically higher F_{IS} and lower mean relatedness values which, in light of the predicted high proportion of allochthonous larval recruits, are attributed to increased variances in reproductive success linked to post-settlement processes. Likewise, microsatellite variation for *C. pagurus* supported a model of a single panmictic population within the Irish Sea. Integration of data collected throughout the Northeast Atlantic indicated high gene flow throughout the studied region but with some level of chaotic genetic patchiness likely due to sweepstakes recruitment episodes. As a comparison, the potential of Restriction-site Associated DNA sequencing (RADseq) as a future tool for population genetic analysis and fisheries management of *C. pagurus* was also investigated, with this being, to our knowledge, the first application of these methods to a crustacean. Seven RADseq libraries permitted genotyping of 566 polymorphic Single Nucleotide Polymorphisms (SNPs) among *C. pagurus* sampled throughout the Northeast Atlantic. Multiple marker-based neutrality tests revealed three consensus positive outlier SNPs. However, these were not significant in pairwise outlier tests and geographical patterns of allelic variation did not lend themselves to robust inference of environmental drivers. The neutral SNP data revealed a lack of wide scale geographic structure but more pronounced chaotic genetic patchiness than reported for microsatellites, indicative of greater sensitivity to neutral structuring. The implications of the findings for marine community ecology, fisheries management and NTZ design strategies are discussed.

ACKNOWLEDGEMENTS

First and foremost, I thank IBERS for the award of a PhD scholarship and SUSFISH - a 3 year project funded by the European Union Regional Development Fund (ERDF) under the Ireland Wales Programme 2007-2013 - Interreg 4A, Project No. 042 - for project consumables and fieldwork allowances. My gratitude goes to my supervisor, Dr Joseph Ironside, for enabling me to have this opportunity.

I am also grateful to all of the IBERS postgraduates and staff members whom I have encountered during my time at Aberystwyth University, for listening to my incessant ramblings and generally making my experience here an unforgettable one. Particular thanks go to Dr Helen Marshall, the tearoom attendees and my office workmates. Additionally, I acknowledge the Gogerddan sequencing facility staff for training and sequencing services.

Special thanks go to my SUSFISH colleagues, especially Dr Emma Wootton (Swansea University) for providing crab and lobster haemolymph samples. Also, I am grateful to Charlotte Davies and Amanda Smith (Swansea University) for the provision of samples. John Hickey (BIM, Ireland) was invaluable during the first sampling trip to Ireland by introducing me to the fishermen whom I would come to depend on for my second sampling trip. These fishermen are: Alan Fanning (Howth), Ivan Toole (Dun Laoghaire), Fintan Parle (Carne) and Martyn Simpson (Dunmore East), plus all of their shipmates. Thanks also go to the fishermen I encountered during my fieldwork in Wales; Dean Parry (Aberystwyth) and Sion Williams (Llŷn Peninsula).

I owe my sincere gratitude to Ilaria Coscia and Niall McKeown for their support and guidance during my PhD, particularly in the final stages. Ilaria mentored me during the three years she was at Aberystwyth University and continued to do so when relocated in Belgium; I am extremely thankful. Niall has been exceptionally helpful in the last few months. I admire his enthusiasm.

Finally, and most importantly, I must thank my family and Geraint for enduring my highs and lows over the last four years. You have been my rock. To my mum, dad and brother - you have always been at the end of the phone when I needed reassurance or someone to just listen. Geraint - you have provided stability at a time of great instability - thank you. Lastly, I must thank my supportive friends, especially Rachel Tate for the surprise PhD pick-me-ups.

Thank you all. Diolch.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Meaning
3D	3-Dimensional
AFLP	Amplified Fragment Length Polymorphism
A_R	Allelic Richness
BR	Broad Range
CI	Confidence Interval
CL	Carapace Length
CPUE	Catch Per Unit Effort
CW	Carapace Width
DNA	Deoxyribonucleic Acid
ENA	Excluding Null Alleles
FDR	False Discovery Rate
H_e	Expected Heterozygosity
H_o	Observed Heterozygosity
H_s	Gene Diversity
HWE	Hardy-Weinberg Equilibrium
IAM	Infinite Allele Model
IBD	Isolation By Distance
MCMC	Markov Chain Monte Carlo
MCZ	Marine Conservation Zone
MLS	Minimum Landing Size
MNR	Marine Nature Reserve
MPA	Marine Protected Area
mtDNA	Mitochondrial DNA

N_A	Number of Alleles
N_b	Effective Number of Breeders
N_c	Census Population Size
N_e	Effective Population Size
NGS	Next Generation Sequencing
NTZ	No-Take Zone
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PLD	Pelagic Larval Duration
POM	Princeton Ocean Model
PTM	Particle Tracking Model
RAD	Restriction-site Associated DNA
RADseq	Restriction-site Associated DNA Sequencing
RFLP	Restriction Fragment Length Polymorphism
RZ	Refuge Zone
SMM	Stepwise Mutation Model
SNP	Single Nucleotide Polymorphism
SOM	Size at the Onset of Sexual Maturity
TAC	Total Allowable Catch
TPM	Two-Phase Mutation Model
UK	United Kingdom

1. GENERAL INTRODUCTION

1.1. Population Genetics and its Application to Fisheries Science

Genetic variation is structured by the action of the evolutionary processes of gene flow, genetic drift, mutation and natural selection. Such processes are influenced on various spatial and temporal scales by the interplay between species life history and environmental processes. Population genetics is the study of genetic variation within and between populations to test hypotheses pertaining to spatial and/or temporal demography. The field of population genetics has a wide range of applications and population genetic approaches have been widely adopted by fisheries science (Carvalho & Hauser 1994; Palumbi 2003; Cadrin et al. 2014). This has seen the utilisation of numerous genetic markers to investigate marine population structure. Each marker type provides a different insight into the spatial and temporal distribution of genetic diversity, with each displaying characteristic differences with regards to transmission and evolutionary dynamics. Nuclear markers such as allozymes, amplified fragment length polymorphisms (AFLPs) and microsatellites are biparentally inherited, whereas mitochondrial DNA (mtDNA) markers are normally maternally inherited. Furthermore, the analysis of restriction fragment length polymorphisms (RFLPs), single nucleotide polymorphisms (SNPs) and direct sequencing can also be used to detect DNA polymorphism.

Genetic markers enable the exploration of population structuring in marine species that might otherwise be undetected due to difficulties in implementing standard ecological methods such as mark-recapture or behavioural observation (Shaklee & Bentzen 1998). Additionally, genetic markers provide a unique ability to distinguish between non-reproductive dispersal and effective dispersal (i.e. those individuals that survive and breed in the new population). Populations that are not linked by effective dispersal (i.e. gene flow) may accumulate different gene frequencies. Therefore, by characterising the geographical distribution of genetic variation population units can be identified. Recent developments of increasingly sensitive genetic markers and associated statistical methods now permit hypothesis testing

at the individual level as well as populations. Such estimates can be used to track the movements of individuals and more precisely quantify gene flow among populations with obvious applicability to stock identification and discrimination (Glover et al. 2011; Beacham et al. 2012).

Populations constitute interbreeding units with more or less autonomous dynamics and recruitment. In terrestrial and freshwater environments, populations are often well delimited by conspicuous physical barriers to mixing and interbreeding (Avise 2000). However, in the marine environment distinct populations are more difficult to detect, with it often being unclear to what degree distinct populations exist at all. The distinction between panmixia and discrete populations is critical, in particular for exploited species, as different populations may possess diverse genetic, physiological or behavioural characteristics that may cause differences in life history traits. As recruitment and sustainability may be properties specific to individual populations, failure to identify and independently manage distinct populations can lead to local overfishing and ultimately to severe declines or stock collapse (Hutchings 2000; Knutsen et al. 2003).

1.2. Population Structure in Marine Species

Many marine species exhibit 'classical' marine traits (Nielsen & Kenchington 2001), such as high dispersal potential, high fecundity, large population sizes and wide geographic distributions, which in conjunction with the lack of conspicuous barriers to dispersal in the marine environment (Briggs 1974), traditionally lead to the assumption of high gene flow (Graves 1998; Bohonak 1999). However, studies reporting significant genetic population structuring for a number of marine taxa have revealed biotic and/or abiotic restrictions to gene flow effective over various spatial and temporal scales (Shaw et al. 1999; Swearer et al. 1999; Jørgensen et al. 2005). Local adaptations have also become increasingly documented among marine species, even against backgrounds of high gene flow, highlighting selection as a potent evolutionary force in the marine environment (Pampoulie et al. 2006; Hemmer-Hansen et al. 2007; Larsen et al. 2007). The prediction that drift is negligible due to large population sizes has also been challenged by empirical

studies reporting effective population sizes to be several orders of magnitude lower than census population sizes (Hauser et al. 2002; Turner et al. 2002). Consequently, genetic studies have driven a paradigm shift in the view of processes shaping marine biodiversity on evolutionary and ecological timescales and revealed extensive intraspecific biocomplexity (Hauser & Carvalho 2008).

Knowledge of the genetic structure of commercially exploited marine species is key to effective management (Thorpe et al. 2000), as it can be used to predict whether a locally depleted stock will be successfully replenished by immigrants (Utter 1991). As a result, there has been considerable research interest in the population genetics analysis of commercially exploited marine taxa. However, the genetic population structure of exploited marine invertebrates has received relatively little attention in comparison to other marine taxa (Weetman et al. 2007). The growing importance of marine invertebrates as a global food resource and associated increases in levels of exploitation (Molfese et al. 2014) has also generated an urgent need to understand patterns of genetic population structure in this diverse group. Several studies report that, even though many marine invertebrates exhibit the aforementioned 'classical' features, significant population differentiation has been observed representing a deviation from the classical assumption of panmixia.

Larval dispersal has been a major focus of investigations of connectivity in marine systems (Becker et al. 2007; Jones et al. 2009; Shanks 2009), however, while several reviews have reported significant relationships between larval dispersal potential and population structuring (Bohonak 1999; Siegel et al. 2003; Kinlan et al. 2005) there are reports wherein dispersal potential poorly predicts realised gene flow patterns (Todd et al. 1998; Fauvelot & Planes 2002; Taylor & Hellberg 2003). More recent meta-analyses have also reported a weaker relationship between larval dispersal potential and population structuring than previously assumed (Bradbury et al. 2008; Ross et al. 2009; Shanks 2009). Such studies have revealed that the genetic structure of populations may be determined by a complex amalgamation of historical demographic signatures (Marko 2004; Hart & Marko 2010) and recurrent interactions between processes that are intrinsic (e.g. larval behaviour; homing),

and extrinsic to species (e.g. water circulation patterns), rather than solely the length of time larvae spend developing in the plankton.

For example, allozyme analysis of two species of spider crab, *Inachus dorsettensis* and *Hyas coarctatus,* indicated significant genetic differentiation between populations only 40 km apart near the Isle of Man, despite both species having a planktotrophic larval phase of several weeks, which is assumed to cause widespread larval dispersal and consequent genetic homogeneity throughout their range (Weber et al. 2000). This unexpected result could potentially be explained by population discontinuities and patterns of larval drift, since population differences were first highlighted by earlier work that revealed differences in growth patterns between neighbouring populations (Hartnoll et al. 1993).

Furthermore, population differentiation in the coastal shrimp, *Crangon crangon*, was investigated with AFLPs (Weetman et al. 2007). *C. crangon* is an ecologically and commercially important species with very little known about its genetic population structure. Three main groupings that corresponded to the geographical regions of western Britain, the eastern English Channel and the Baltic Sea were revealed, with significant differentiation identified both within and among these regions. This species demonstrates greater genetic structuring than has been observed in many decapod crustaceans studied to date, with the findings appearing consistent with unstable population sizes, gene flow restricted by distance, and hydrographic features. It is thought that hydrographic barriers could act as important determinants of genetic differentiation in *C. crangon*, potentially superseding the effect of distance on dispersal limitation. Therefore, despite the high dispersal potential, homogeneity of stocks should not be assumed.

Unfortunately, the amount of published genetic studies on the effects of pelagic larvae on gene flow in exploited marine invertebrates is somewhat lacking; several studies have indicated that pelagic larvae mediate widespread gene flow with little or no genetic differentiation over large distances (Mulley & Latter 1981; Silberman & Walsh 1994), yet the aforementioned studies (Weber et al. 2000; Weetman et al. 2007) argue this. It is clear that more work needs to be done to assess the population structuring of commercially important species in areas of exploitation.

1.3. Commercial Decapods in the Irish Sea

1.3.1. *European Lobster*

European lobster, *Homarus gammarus*, is present in coastal waters in the Northeast Atlantic, ranging from north of the Arctic Circle in Norway to Morocco; primarily occurring from the low intertidal to depths greater than 50 metres (Wahle et al. 2013). Total body length can exceed 400 mm, with specimens found intertidally generally much smaller (Fish & Fish 2011). The carapace is dark blue in colour and the first pair of pereopods terminates in large chelae that are unequal in size (Figure 1.1).

Size at the onset of sexual maturity (SOM) of *H. gammarus* varies geographically due to the dependence on summer water temperature (Wahle et al. 2013). For instance, around the coasts of the United Kingdom (UK) and Ireland, the median size at which female *H. gammarus* bear eggs ranges from 80 to 100 mm carapace length (CL) (Free et al. 1992; Tully et al. 2001; Lizárraga-Cubedo et al. 2003; Laurans et al. 2009; Wahle et al. 2013). In most areas, lobsters do not mature before five to eight years of age (Prodöhl et al. 2007). Growth is by moult, which decreases in frequency during the early life stages until becoming part of the annual mating, spawning and egg hatching cycle (Factor 1995). The mating system in *H. gammarus* is polygynous, with the larger, competitively dominant males acquiring the prime mating shelters (Wahle et al. 2013). Mating occurs during summer and is linked with the moulting cycle (Atema 1986). Once extruded, eggs are held on the pleopods of the female for about nine months. The capacity of sperm storage in *Homarus* has been suggested from the ability of larger females to moult and endure two consecutive spawns prior to moulting again (Waddy & Aiken 1986). When hatched, the larvae remain planktonic for three weeks, progressing through three larval instars and one postlarval instar, before settlement on the sea bed by the late-stage post-larva (Wahle et al. 2013). Early juveniles are cryptic and remain strongly associated with shelter, such as coarse sediments, in which they construct complex burrows (Howard & Bennett 1979).

Figure 1.1. The European lobster, *Homarus gammarus*, being V-notched.

This decapod is long-lived, capable of surviving more than several decades (Sheehy et al. 1999), although fishing pressure severely reduces the chance of many reaching those later years. *H. gammarus* are sedentary animals with small home ranges, varying from 2 to 10 km (Prodöhl et al. 2007), thus the pelagic larval phase is considered as the main mechanism of dispersal.

1.3.2. *Edible Crab*

Edible or brown crab, *Cancer pagurus*, is widely distributed in the Northeast Atlantic from northern Norway to northwest Africa (Christiansen 1969), with evidence of its biogeographical range expanding further northwards (Woll et al. 2006). The carapace is much broader than long (< 200 mm across) and is reddish-brown in colour with a distinctive 'pie-crust' effect (Figure 1.2).

C. pagurus mature at 101.6-125.0 mm (male) and 110.7-147.3 mm (female) carapace width (CW) (Tallack 2007b), when they are approximately three to five years old. Copulation can only occur once the female has moulted and is in a soft bodied state (Edwards 1966; Hartnoll 1969; Edwards 1979). Consequently, the male will attend the female for periods of 3 to 21 days prior to the female moulting (Edwards 1966). Further to this precopulatory behaviour, mate guarding will continue for 1 to 12 days after mating (Edwards 1966; Edwards 1979), which is

commonly seen in species with soft-shelled female mating (Hartnoll 1969). Insemination occurs and the sperm is stored until spawning; the interval between mating and spawning is highly variable, ranging from a few months to up to 15 months (Pearson 1908). Despite the ability of long-term sperm storage, single paternity is the predominant system in *C. pagurus* (McKeown & Shaw 2008b). Breeding takes place in winter and the eggs hatch in the spring or summer, following a seven to eight month period of incubation. The larvae are planktotrophic and remain pelagic for approximately three months when water temperatures are 14 °C ± 3 °C (Eaton et al. 2003; Weiss et al. 2009; Hunter et al. 2013), advancing through five planktotrophic zoeal stages and a megalopa before reaching the first crab stage (Ingle 1981).

Figure 1.2. The edible or brown crab, *Cancer pagurus*.

Adult *C. pagurus* are characterised as mobile and benthic, although sex-specific differences in migration have been found; females tend to migrate significantly longer distances more frequently than males (Edwards 1979; Bennett & Brown 1983; Latrouite & Le Foll 1989; Ungfors et al. 2007). Size differences in migration have also been detected, with the larger males and females tending to make the more extensive movements (Bennett & Brown 1983). Adult migrations are linked to

breeding behaviour, with males migrating to ensure suitable-sized mates for copulation when the females moult and are sexually receptive, whilst berried females are assumed to migrate to locate suitable substrate (Bennett & Brown 1983). Berried females are not readily caught in baited pots, as they do not feed during most of the six to nine months prior to their eggs hatching (Howard 1982). This accounts for the lower catch rates of female crabs after November, when they have become berried (Brown & Bennett 1980).

C. pagurus live for up to 20 years, with females estimated to carry 0.5 to 2.9 million eggs at any one time (Edwards 1979; Ungfors 2007), resulting in a highly fecund species with great dispersal potential.

1.4. Shellfisheries and their Management

In 2013, the UK fishing industry had 6,399 fishing vessels, comprised of 5,036 10 metre and under vessels and 1,363 over 10 metre vessels. These statistics show a declining trend, since 7,096 vessels were in the UK fishing fleet ten years earlier; a 10% reduction compared with 2003 (Marine Management Organisation 2014). This industry provides employment for 12,150 fishermen, with around 5,600 of these based in England, 5,000 in Scotland, 810 in Northern Ireland and 730 in Wales (Marine Management Organisation 2014). Part-time fishermen account for only 15% of the total, thus highlighting the socio-economic importance of the industry. In 2013, UK fishing vessels landed 624,000 tonnes of sea fish (including shellfish) in the UK and abroad with a value of £718 million. Shellfish accounted for 37% of the total landings by value, with demersal and pelagic fish accounting for 38% and 25% respectively (Marine Management Organisation 2014).

In Wales, shellfish are the main focus of the fleet, with 69,000 tonnes landed in 2013, compared to only 700 tonnes of finfish. Shellfish is a valuable commodity, as the average value of shellfish landed by UK vessels into the UK is £1,743 per tonne (live weight), compared with £1,658 per tonne for demersal species and £658 per tonne for pelagic species (Marine Management Organisation 2014). The three main species of shellfish are scallops, crabs and nephrops (langoustines), accounting for

72% of the quantity and 73% of the value landed in 2013 by UK vessels into the UK and abroad. In addition, lobsters are highly valuable and command the highest average price of all species landed by the UK fleet, with the 3,000 tonnes landed worth £29.8 million. Crabs are more heavily exploited than lobsters and are valued much lower, with the 28,800 tonnes landed worth £38.5 million (Marine Management Organisation 2014).

In the UK, the imports of shellfish greatly outweigh the exports, with 117,300 tonnes worth £676.1 million imported and 88,600 tonnes worth £451.9 exported. Crabs and lobsters constitute 2,500 and 2,600 tonnes of the imports, worth £17.3 and £23.4 million, respectively. However, export values for these two species are much higher, with 14,300 tonnes of crabs worth £50.8 million and 7,400 tonnes of lobsters worth £75 million (Marine Management Organisation 2014).

Evidently, shellfish are highly valuable to the UK fishing industry and, therefore, should be managed accordingly so as to prevent stock collapse. Several fishery restrictions are in place to manage shellfisheries around the UK and those for *H. gammarus* and *C. pagurus* are described herein.

In Europe, lobsters and crabs are managed at EU, national, regional, and local level. The main technical measure is the enforcement of a minimum landing size (MLS), which is designed to ensure that the shellfish are able to grow to maturity to sustain breeding stocks (SEAFISH 2013). When the MLS is not met, the undersized animal is returned to the sea where it can continue to contribute to the population. The EU MLS for *H. gammarus* is 87 mm CL (SEAFISH 2013), with national legislation for the crustacean fishery in Wales increasing this to 90 mm in the south region (0-6 nautical miles) (Welsh Government 2014). For *C. pagurus*, the EU MLS varies around Britain, reflecting the differences in growth rate and size at first maturity throughout its range. In the Irish Sea, the MLS in effect is 130 mm CW (SEAFISH 2013). Similarly, national legislation has set the MLS of *C. pagurus* in the south region of Wales to 140 mm (Welsh Government 2014). Furthermore, it is prohibited to land detached parts of crustaceans, including *H. gammarus* and *C. pagurus*, in the south region, with *H. gammarus* demanded to be landed whole in all parts of Wales (Welsh Government 2014).

A further technical measure that some fishermen in the UK and Ireland carry out is the notching of a V-mark into the tails of breeding lobsters (Figure 1.1) prior to releasing them. Whilst the act of V-notching is voluntary, once V-notched it is illegal for that lobster to be landed. A V-notch is a successful way of denoting breeding stock all year round, even when the eggs are not visible (SEAFISH 2013).

Another management measure is the shellfish licencing scheme, which all UK pot fisheries for lobsters and crabs are controlled through. This scheme restricts entry of new vessels to the fishery and demands catch and fishing effort information. In addition, EU or national legislation on crabs prohibits landing berried females and soft pre-moult or recently moulted crabs. At present, *H. gammarus* and *C. pagurus* are not managed through a Total Allowable Catch (TAC) or quota allocation.

1.5. SUSFISH Project

This work formed part of the EU Wales-Ireland Interreg project SUSFISH (Shellfish productivity in the Irish Sea: working towards a sustainable future) (www.susfish.com). SUSFISH is a consortium project involving Aberystwyth, Bangor, Cork and Swansea Universities, unifying experts from a wide range of disciplines, including shellfish biology, ecology, physical oceanography, and economics. The collaborative aim was to produce guidelines for future fisheries management and policy of the shellfish industry in Ireland and Wales for the next 50-100 years; achieved by assessing the biological, genetic, environmental and economic impacts of climate change on shellfish productivity in the Irish Sea, and determining adaptation or mitigation strategies for the shellfish industry.

Climate change is a very real cross-border threat and the link between increased incidence of disease and mortality in marine organisms to climatic alterations and anthropogenic factors is becoming more evident (Harvell et al. 1999; Harvell et al. 2002; Hoegh-Guldberg & Bruno 2010). The changing climate is a concern for industries that rely upon marine resources, particularly those in rural areas whose economies would be significantly altered. SUSFISH focused on the Irish Sea, which provides for shellfisheries in Ireland and Wales; a major economic resource with

total shellfish landings in 2002 (52,500 t) worth approximately 62.2 million euros (Malham 2010). The project intended to produce positive opportunities for the Irish Sea shellfish industry from the possible negative effects of global climate change. A suite of techniques were used to identify the effects that climate change will have on shellfish stocks, including oceanographic models linked to climatic data, historic and current data collection, experimental research, and the use of economic assessments to identify economic benefit and sustainable development of the region.

Ten work packages were created to achieve the desired output, with Aberystwyth University responsible for the population genetics work package GENEFISH. Its aim was to assess the population genetic structure and mixing of the common cockle (*Cerastoderma edule*) and the edible crab (*Cancer pagurus*), two of the most commercially important shellfish species in the Irish Sea. GENEFISH results were used to test the predictions of the larval dispersal models and to infer losses of genetic diversity due to overexploitation and/or mass mortality events; see Coscia et al. (2013) for the cockle research outcomes. The results of this work package will enhance the management of crab and cockle populations by indicating the spatial scale at which discrete, self-recruiting stocks may exist and providing baseline data for future genetic monitoring schemes.

SUSFISH was a three year project funded by the European Regional Development Fund (ERDF) through the Ireland Wales Programme 2007-2013.

1.6. Outline of Thesis

This research focused on the population genetic analysis of two socio-economically important crustacean species, the edible crab (*Cancer pagurus*) and European lobster (*Homarus gammarus*), with emphasis on investigating potential population structuring within the Irish Sea.

In Chapter 2, genetic data collected for *H. gammarus* was interpreted in conjunction with biophysical modelling of larval dispersal to assess (i) the relationship between predictions of larval connectivity and realised gene flow, and (ii) the genetic patterns associated with a No-Take Zone (NTZ). A similar microsatellite-based genetic analysis of *C. pagurus* in the Irish Sea was performed in Chapter 3, also including samples from the NTZ. Together, these chapters provide a comparison of the genetic structure for two species with distinct ontogenetic movement patterns within the Irish Sea. Genetic diversity of *C. pagurus* was then placed within a wider geographic context, achieved by the addition of genotypic data from samples collected from throughout the Northeast Atlantic. In line with the growing trend of applying population genomic methods to study marine biocomplexity, a preliminary investigation into the utility of Restriction-site Associated DNA Sequencing (RADseq) to identify neutral and adaptive population structuring in *C. pagurus* is reported in Chapter 4. The final chapter (Chapter 5) provides an overview of the results of this research, with discussions on the implications for future management, conservation and studies of marine taxa.

To conclude, the questions addressed in this PhD thesis are:

1. Is there significant genetic structuring in *H. gammarus* and *C. pagurus* in the Irish Sea, or do they constitute single biological stocks in each case?
2. Do *H. gammarus* and *C. pagurus* within the Lundy NTZ exhibit distinct patterns of genetic diversity, such as differing levels of genetic variation, relatedness and allele frequency distributions, compared to fished areas?
3. Does RADseq confer increased resolution of neutral and adaptive structuring of *C. pagurus* populations?
4. What are the implications of these findings for the conservation and management of these and other commercially harvested decapods?

2. POPULATION GENETIC STRUCTURE OF THE EUROPEAN LOBSTER, *HOMARUS GAMMARUS*, IN THE IRISH SEA: COMPARING REALISED DISPERSAL WITH PREDICTIONS FROM A LARVAL DISPERSAL MODEL

2.1. INTRODUCTION

The majority of marine species have a dispersive larval phase in their life history, with time spent in the plankton ranging from hours to months (Pechenik 1999). For species with limited adult movements, this short larval stage denotes the main dispersal opportunity (Cowen et al. 2006). Whilst population connectivity is driven by the dispersal of such individuals (Palumbi 2003), the scale at which it is occurring and the influence of ocean currents remains largely unknown. For the most part, this is due to microscopic larvae (~200 µm) being extremely difficult to study in open marine environments (~km) (Gilg & Hilbish 2003). Consequently, attention has turned to biophysical models to answer the fundamental questions of where do larvae come from (i.e. connectivity) and where do they go (i.e. dispersal) (Levin 2006).

Dispersal plays a significant role in the structuring and connectivity of marine populations and has been the focus of many studies investigating the link between pelagic larval duration (PLD) and population structure. In marine organisms, it is generally thought that increasing PLD should result in increasing dispersal distance and, therefore, higher connectivity/gene flow and lower genetic structure among populations (Palumbi 1992; Bohonak 1999). However, several studies have challenged this paradigm by identifying high levels of genetic structure in species with long-lived planktonic larvae (Swearer et al. 1999; Barber et al. 2000; Taylor & Hellberg 2003). Furthermore, weak or no correlation between PLD and F_{ST} has been found in more recent studies (Kelly & Palumbi 2010; Riginos et al. 2011), suggesting that PLD may not be a good predictor for population connectivity (Shanks 2009; Weersing & Toonen 2009; Selkoe & Toonen 2011). Reasons for such deviations include: retention caused by local oceanographic features (Teske et al. 2007), larval behaviour (Imron et al. 2007), and error inherent in estimating PLD and F_{ST} (Faurby

& Barber 2012). Therefore, it seems that PLD alone may inaccurately predict population genetic structure in marine invertebrates.

In crustaceans, a number of studies have shown that genetic structure is not correlated with dispersal potential (*Nephrops norvegicus*, Stamatis et al. 2006; *Crangon crangon*, Weetman et al. 2007). For instance, significant genetic differentiation was revealed between populations of spider crabs, *Inachus dorsettensis* and *Hyas coarctatus*, over a geographical distance of only 40 km, even though both species have a life cycle with an obligatory planktonic larval phase of several weeks (Weber et al. 2000). Iacchei et al. reported localised self- recruitment and/or larval cohesion in the California spiny lobster, *Panulirus interruptus*, a species with a 230-330 day PLD. In addition, Kennington et al. (2013) describe fine scale ephemeral chaotic genetic patchiness occurring against a background of high gene flow in the western rock lobster, *Panulirus cygnus*, indicative of restricted movements and/or spatial cohesion of individuals after larval settlement. For commercially exploited species, failure to identify patterns and processes underlying population structure may result in overexploitation and depletion of local stocks with a corresponding loss of biodiversity (Carvalho & Hauser 1994).

The European lobster, *Homarus gammarus*, is widely distributed in the Northeast Atlantic, with its range extending from the Arctic Circle to Morocco, although it is not present in the Baltic Sea due to lowered salinity and temperature extremes (Triantafyllidis et al. 2005). A paucity of information on wild *H. gammarus* larvae has seen data from laboratory observations and from the closely-related and extensively-studied American lobster, *H. americanus*, being referred to. Clawed lobsters have four planktonic life stages, three larval (stages I-III) and one postlarval (stage IV), with each stage potentially exhibiting different behaviours; stages I-III are capable of slow swimming in the vertical plane and are able to maintain or change vertical position in the water column, whilst stage IV are capable of rapid forward swimming and diving, and cluster at the surface (Cobb & Wahle 1994). In addition, *H. gammarus* larvae make a diel vertical migration toward the surface at dusk and dawn (Nichols & Lovewell 1987). Settlement occurs during stage IV, in which the larvae aim to find and select suitable benthic habitat. Larval stages I, II and III have

durations of 5-8, 5-10, 6-10 days, respectively, with stage IV starting after 16-28 days (Ouellet & Allard 2002).

Populations genetics of *H. gammarus* was studied in depth as part of the GEL-FAIR (Genetics of the European Lobster) project using a combination of molecular markers (Prodöhl et al. 2007). Four distinct groups were identified throughout the distribution range of *H. gammarus* with microsatellite markers (Ferguson 2002), mitochondrial DNA (mtDNA) (Triantafyllidis et al. 2005), and allozymes (Jørstad et al. 2005): northern Norway, Netherlands, remaining Atlantic samples, and the Mediterranean, particularly the Aegean. The northern Norway, Netherlands and Aegean groups differentiate from the main Atlantic group due to reduced gene diversity and not because of unique alleles (Prodöhl et al. 2007). Such low levels of genetic diversity point towards the existence of limited gene interchange among them and a recent establishment from a common refugium after the end of the last Ice Age, within the past 15,000 years (Ferguson 2002; Triantafyllidis et al. 2005). The population structure found along with the limited gene flow means that local overexploitation would not be counter-acted by recruitment from elsewhere within time periods of relevance to management. Consequently, management and conservation should target local populations and not the *H. gammarus* metapopulation as a whole (Ferguson 2002).

In recent years, one approach of dealing with overexploitation of marine resources through heavy fishing activity has been the establishment of Marine Protected Areas (MPAs) or No-Take Zones (NTZs) (Hoskin et al. 2011; Moland et al. 2013; Øresland & Ulmestrand 2013). The species afforded the most protection in marine reserves are those with a limited adult home range. Adult *H. gammarus* are sedentary with small home ranges and strong site fidelity (Jensen et al. 1993; Bannister et al. 1994; Smith et al. 2001; Moland et al. 2011; Øresland & Ulmestrand 2013), thus the pelagic larval phase is predicted to be the main mechanism of dispersal. From 1993 to 2007, there was a marked increase in the abundance of *H. gammarus* in the Swedish Kåvra lobster reserve, with the number of lobsters caught per trap per day rising by 252% (Øresland & Ulmestrand 2013). Also, within four years (2006 – 2010) of MPA establishment along the Norwegian Skagerrak coast,

there was a 245% increase in the catch-per-unit-effort (CPUE) of *H. gammarus*, as well as a 13% increase in mean size (Moland et al. 2013).

In the UK, lobsters are an economically important species, representing more than 10% of the total value of shellfish landings (£31.099 million out of £300.801 million) whilst only comprising less than 2% of the shellfish landed (3,159 tonnes out of 162,754 tonnes) (Cefas 2013). In 2009, the UK's first Marine Conservation Zone (MCZ) was granted by the UK Marine and Coastal Access Act at Lundy Island, located in the Bristol Channel off the southern Welsh coast. It is approximately 5 km by 1.25 km in size and, prior to this, was designated as a Marine Nature Reserve (MNR) in 1986, which included a Refuge Zone (RZ) extending 1.5 km offshore, before its NTZ designation in 2003 (3.3 km^2). Both the RZ and the NTZ are maintained within the MCZ status. Consequently, the Lundy NTZ provides an opportunity to test hypotheses about the recovery of crustacean populations from fishing. Within four years of protection, there was a 127% increase in the abundance of legal-sized *H. gammarus* observed in the Lundy Island NTZ, making them five times more abundant within the protected waters compared to regions outside (Hoskin et al. 2011). Evidently, MPAs could be an effective management tool, although the implications for wider population dynamics remain unanswered.

Modern genetic approaches can be a reliable support by providing empirical assessment of connectivity among marine populations (Hellberg et al. 2002; Hedgecock et al. 2007; Lowe & Allendorf 2010). Incorporating them with statistical modelling of larval dispersal, integrating both biological and physical parameters, represents a powerful tool for predicting population patterns under different conditions (James et al. 2002; Cowen et al. 2006; North et al. 2008), especially when sampling early life stages (larvae) in the wild is extremely difficult, like in the case of the European lobster. Integrated population genetic and biophysical modelling thus offers considerable potential for the elucidation of current population structuring patterns and mechanisms (White et al. 2010; Coscia et al. 2013), as well as predicting the future changes they might undergo in a global change scenario.

The primary objective of this study was to describe the population genetic structure of *H. gammarus* in the Irish Sea using a combination of traditional pairwise and

global tests of genetic diversity, and Bayesian clustering. As population genetic structure reflects realised connectivity (i.e. interbreeding), a secondary aim was to compare patterns of genetic connectivity with estimates of larval connectivity derived from a biophysical model of larval dispersal. *H. gammarus* larvae dispersal from the Lundy NTZ was simulated using coupled hydrodynamic and particle tracking models, in addition to integrating behavioural strategies (Wootton et al. in prep). To address this, standard inter-sample comparisons of allele frequencies were combined with coalescent based tests of asymmetric dispersal. Findings from this study will enable the impact of a small marine reserve on fisheries enhancement and sustainability to be assessed.

2.2. MATERIALS AND METHODS

2.2.1. Biophysical Modelling

The biophysical model was developed by SUSFISH colleagues at Bangor and Swansea Universities (Wootton et al. in prep) and is briefly described here.

Simulations of larval transport of *Homarus gammarus* in the southern Irish and Celtic Seas were modelled with a biophysical model comprising two components: a 3-dimensional (3D) hydrodynamic model and a Lagrangian particle tracking model (PTM). Larvae particles were released from the marine reserve at Lundy Island with the aim of predicting dispersal to establish the proportion of self-recruitment and the extent of connectivity of this population with other populations.

The model domain encompasses the Irish and Celtic Seas, extending north to the North Channel (Figure 2.1). The domain is typical of semi-enclosed mesoscale (10 – 1000 km) basins around the world (Taylor 1919) and, therefore, analogous to other marine ecosystems. However, the tidal ranges in the Bristol Channel are extremely large (> 11 m during spring tides), thus generating large tidal flows around Lundy Island and in the Bristol Channel, with a degenerate amphidrome (zero elevations but strong tidal velocities) located off the southeast coast of Ireland (Robinson 1979; Neill et al. 2009). The larval transport simulations focus on larval transport in the shelf sea; transport within estuaries or bays may depend on local conditions and must be addressed separately (Robins et al. 2012). Therefore, larvae are assumed to be located off-shore and will be transported to either their natal habitat (self-recruitment) or to a similar habitat elsewhere (connectivity).

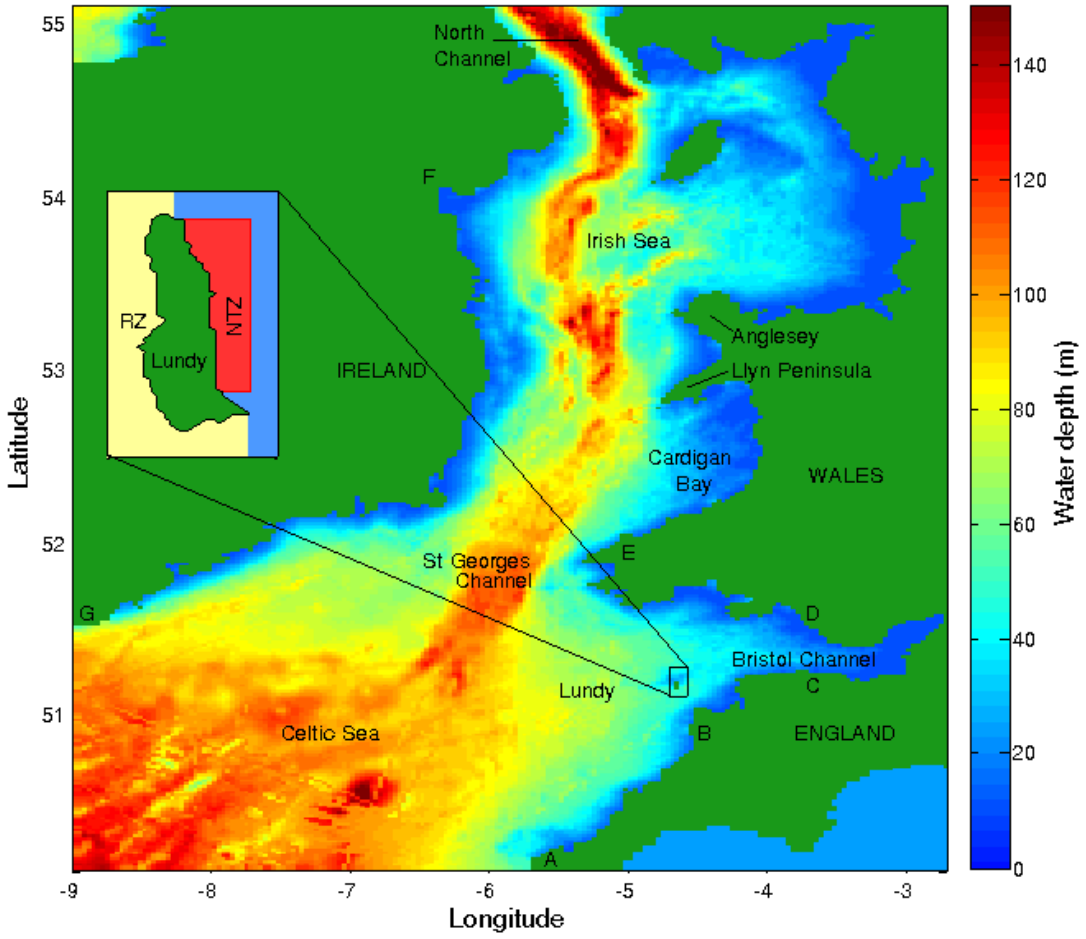


Figure 2.1. The 3D hydrodynamic model domain, showing water depths at mean sea level. The Irish and Celtic Seas are bounded by St George's Channel and the Irish Sea extends north to the North Channel. The horizontal grid resolution is ~ 1.85 km, and there are 20 terrain-following vertical layers. Land is coloured green and the English Channel (in the south-east) has been omitted from the hydrodynamics (water depths here not accurate). Coastlines and bathymetries near open boundaries have been smoothed to minimise instabilities from developing. Mesoscale baroclinic circulations within the Irish and Celtic Sea were accurately resolved. Several coastlines have been highlighted for the analysis: north Cornwall (A-B), north Devon (B-C), Bristol Channel (C-D), south-west Wales (D-E), and Ireland (F-G). The inset map shows Lundy Island (boxed area on main map); the red region (Lundy East) is the marine reserve, or No-Take Zone (NTZ), and the yellow region (Lundy West) is the Refuge Zone (RZ). Modified from Wootton et al. (in prep).

2.2.1.1. *Hydrodynamic Model*

A parallel version of the 3D free-sea-surface Princeton Ocean Model (POM) (Mellor & Yamada 1982; Blumberg & Mellor 1987) was used as the model system for the Irish Sea (Robins et al. 2013). The model horizontal cell is 1/30° (longitude) by 1/60° (latitude), giving a resolution of approximately 1.85 km. In the vertical plane, 20 equally segmented sigma-layers give minimum and average resolutions at mean sea level of 9.6 m and 4.3 m, respectively.

In order to validate the model, a year in the recent past, namely 1990, was chosen as a mean year for the simulation. This was based on output of bed shear stress, significant wave heights and sea surface temperatures from a decadal simulation (1989-1998) of an outer-nested model of the Northwest European Shelf (Neill et al. 2010). The 1990 simulation was used both for the validation process and the hydrodynamic model input for the PTM. Rationale, methodology and results of the validation process are presented in (Robins et al. 2013).

2.2.1.2. *Particle Tracking Model*

Lagrangian PTMs were used in conjunction with the Irish Sea POM (hydrodynamic model) to simulate individual particle displacement in space and time based on advection, sub-grid-scale turbulent mixing, and individual particle behaviour (Robins et al. 2013). Three-dimensional velocity and diffusivity output from the hydrodynamic model are used in the PTMs. The advantage of this method is that both the hydrodynamic model and the PTMs are more computationally efficient when separated and the hydrodynamics can be used to simulate a large number of cluster-release scenarios, without multiple hydrodynamic simulations. Velocity outputs have been tri-linearly interpolated within the PTM to the position of each particle, with the time interval of the hydrodynamic model (15 minutes) linearly interpolated to five minutes for the PTM. Each particle was then iteratively advected in space and time.

In the absence of the known larval migration strategy for *H. gammarus*, simulations were performed for six different larval behavioural strategies (PTM 1-6) (Table 2.2).

Table 2.1. The larval behaviour strategies implemented in the Particle Tracking Model and the Particle Backtracking Model. Modified from Wootton et al. (in prep).

PTM Scenario	Larval Depth or Strategy			
	Stage I: Week 1	Stage II: Week 2	Stage III: Week 3	Stage IV: Weeks 4-8
PTM-1	Passive	Passive	Passive	Passive
PTM-2	2.5 m	2.5 m	2.5 m	1 m
PTM-3	5 m	5 m	5 m	1 m
PTM-4	10 m	10 m	10 m	1 m
PTM-5	15 m	15 m	15 m	1 m
PTM-6	Diel	Diel	Diel	1 m

PTM-1 was a control scenario with no migration strategy, i.e. passive particles for stages I-IV. For PTMs 2-5, larvae were positioned at different depths within the water column (2.5 m, 5 m, 10 m, and 15 m, respectively) for larval life stages I-III, with larvae located in the surface layer (1 m depth) during stage IV. PTM-6 was a diel migration strategy, which simulated larvae travelling to the surface layer during the night and sinking down the water column during the day, with larvae located in the surface layer (1m depth) during stage IV.

For each PTM scenario, cohorts of 10,000 larvae were released within the Lundy NTZ on the east coast of the island. Larvae were released on 16 dates, at 5 day intervals between 15th May and 29th July, simulating 160,000 individual larvae. Each experiment was repeated for all six larval behavioural strategies, totalling 96 PTM simulations and 960,000 larvae trajectories. During the simulations, if land was encountered, the larvae were reflected back into the water column to its position at the previous iteration step, in order to investigate maximum dispersal (North et al. 2008). Likewise, no mortality was assigned to the larvae so as to determine the geographic dispersal capability of larvae from the Lundy NTZ. The location of each particle was recorded at the end of each larval stage, at weeks 1, 2, 3 and 8 after release, with trajectories tracked for a total of 8 weeks.

Upon completion of each PTM simulation, a similar criterion to (Cowen et al. 2006) was used to determine which larvae reached suitable settlement sites; any larvae within 10 km of land at the end of stage IV were considered successful settlers. A distance of 10 km was chosen because it is comparable to the distance that a

passive particle travels in mean currents of 0.5 m s^{-1} (Robins et al. 2013), over a flood or ebb phase of the tidal cycle. Larvae that did not meet these criteria by being located in deeper water were considered unsuccessful and were not included in estimations of connectivity and self-recruitment of Lundy NTZ.

2.2.1.3. *Particle Backtracking Model*

A particle backtracking model was implemented in an attempt to answer the fundamental question of where larvae come from. During backtracking modelling, larvae are released from a known settlement site and their position tracked backwards in time through the velocity field, thus making particle backtracking models a useful tool for estimating an initial probability distribution of a particular population (Batchelder 2006). Cohorts of 10,000 larvae were released from Lundy NTZ and tracked backwards in time for a PLD of 8 weeks, using the same larval behaviour strategies as for the forward tracking models (see Table 2.1), although the larval stages were run in reverse. In order to analyse larval distributions at the end of each larval stage, each simulation was divided into four separate simulations: stages IV-I, III-I, II-I, and I. Source probability distributions for larvae that settle at Lundy NTZ were calculated from instantaneous larval positions of all larvae tracked throughout the season, a total of 160,000 particles, for each behavioural strategy and for each larval life stage duration.

2.2.2. Sample Collection

Haemolymph was collected from a total of 385 lobsters at nine locations within the Irish Sea (Table 2.2, Figure 2.2). Sampling locations were along the coasts of Wales [North Llŷn Peninsula (NW), Aberystwyth (CB) and the Gower (SW)] and Ireland [Howth (ND), Carne (WEX) and Dunmore East (WF)]. Samples were also collected at Lundy Island in the Bristol Channel [Lundy NTZ (LNTZ) and outside the NTZ (LICZ)] and Ilfracombe, North Devon (DEV). Commercially harvested individuals were sampled, therefore sampled individuals are above the MLS of 90 mm carapace length (SEAFISH 2013).

Table 2.2. The location, assigned code and coordinates of the sampling locations of *Homarus gammarus* in the Irish Sea, as well as the date of sample collection (S) and the number of individuals sampled (N). The sex of the individual was recorded when possible (M : F).

Location	Code	Coordinates	S	N	M : F
Howth, North Dublin (IE)	ND	53.469° N, 6.084° W	Jul-12	36	11 : 25
Carne, Wexford (IE)	WEX	52.184° N, 6.302° W	Jul-12	29	12 : 16
Dunmore East, Waterford (IE)	WF	52.085° N, 7.033° W	Jul-12	48	20 : 28
North Llŷn Peninsula, North Wales (UK)	NW	52.806° N, 4.823° W	Oct-12	40	24 : 16
Aberystwyth, Cardigan Bay (UK)	CB	52.415° N, 4.236° W	Jun-12	44	19 : 25
Gower, South Wales (UK)	SW	51.550° N, 4.144° W	Nov-12	48	23 : 25
Lundy Island (outside NTZ) (UK)	LICZ	51.205° N, 4.682° W	Aug-11	44	18 : 26
Lundy Island NTZ (UK)	LNTZ	51.189° N, 4.649° W	Aug-11	48	25 : 23
Ilfracombe, North Devon (UK)	DEV	51.228° N, 4.125° W	Jun-13	48	21 : 27

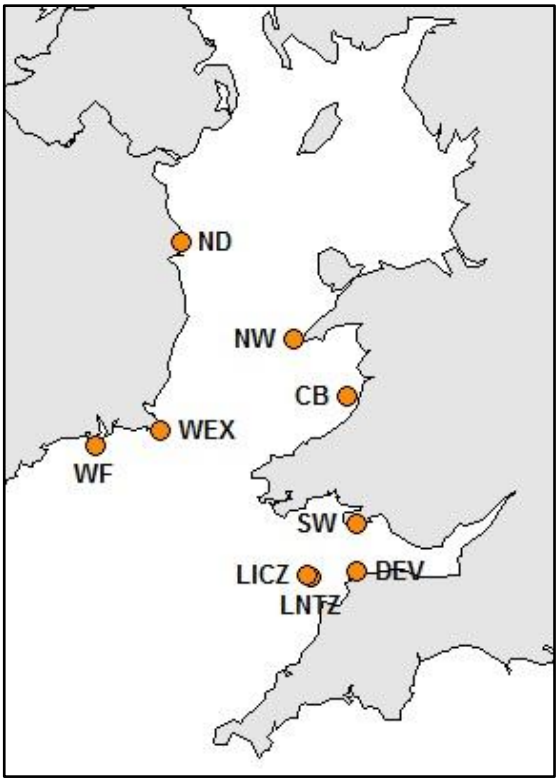


Figure 2.2. Sampling locations of the European lobster *Homarus gammarus* in the Irish Sea.

2.2.3. DNA Extraction and Microsatellite Genotyping

Haemolymph was taken using a 2 ml Terumo syringe with a G23x25 mm needle (VWR International Ltd.) and preserved in absolute ethanol (1:8). DNA was extracted using the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, CA-USA) using the Animal Blood Spin-Column Protocol, which was modified to include an initial centrifugation step of 400 µl of the haemolymph/ethanol mixture (sample) for 5 minutes at 7000 x g in order to precipitate the haemolymph cells into a pellet easily separated from the alcoholic fraction.

Twelve species-specific microsatellite loci developed by André & Knutsen (2010) were amplified in two multiplex PCRs (Figure 2.3). In both multiplex reactions, amplification was carried out using a QIAGEN Multiplex PCR Kit (QIAGEN, CA-USA) in a final volume of 10 µl. This contained 5 µl of Multiplex Kit Buffer, 1 µl of genomic DNA and 0.2 µl of each the forward (fluorescently labelled) and reverse primers for the specific multiplex reaction. The PCR cycle involved an initial denaturation step at 95 °C for 15 minutes, followed by 34 cycles of 45 seconds at 94 °C, 45 seconds at 59 °C and 45 seconds at 72 °C, and a final extension step at 72 °C for 45 minutes. Products were then run on an ABI 3730 Genetic Analyzer (Applied Biosystems) alongside a GS500LIZ size standard and alleles inferred using GeneMapper 4.0 (Applied Biosystems).

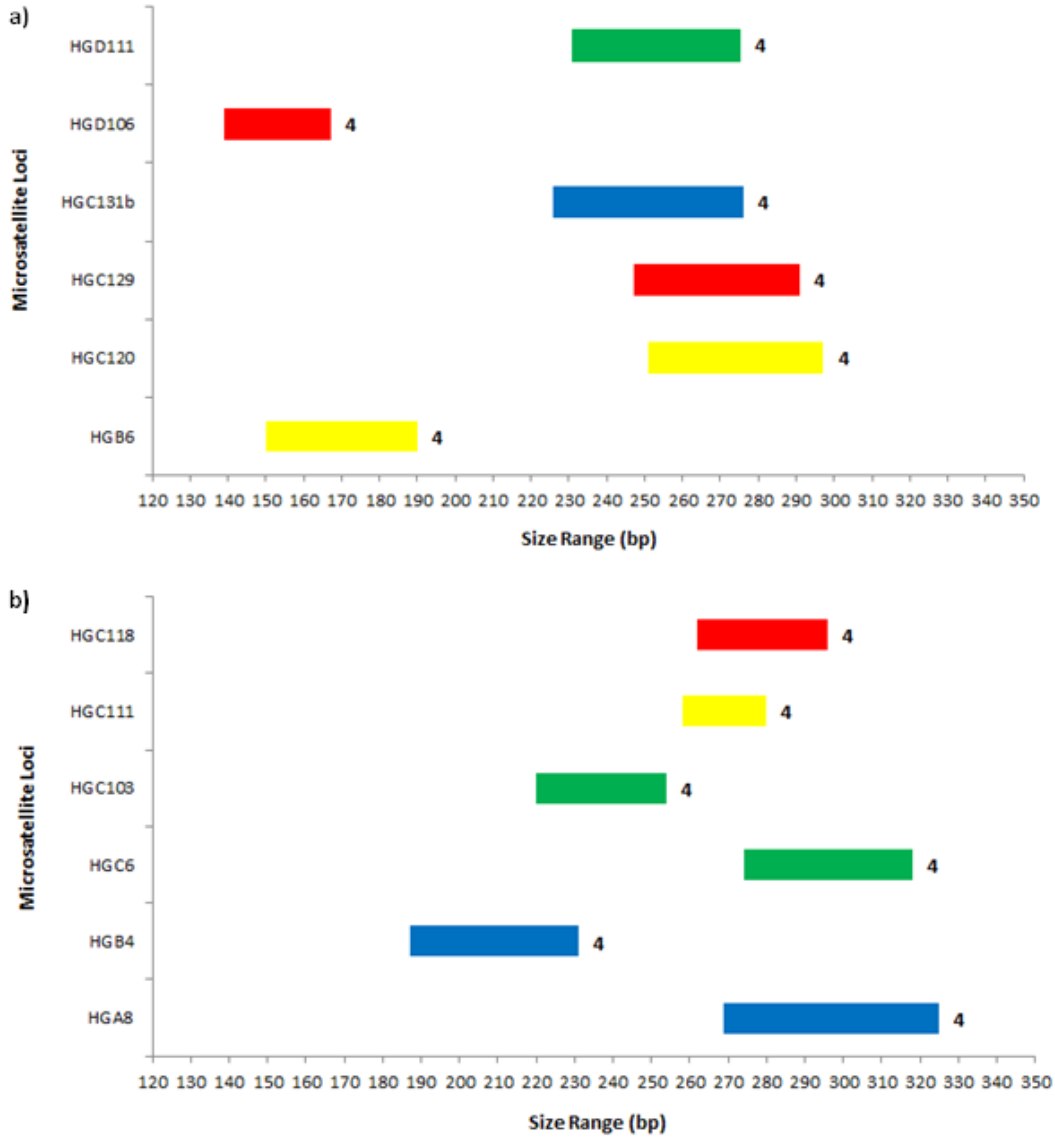


Figure 2.3. (a) Multiplex Plan 1, for six microsatellite loci (HGD111, HGD106, HGC131b, HGC129, HGC120 and HGB6). (b) Multiplex Plan 2, for six microsatellite loci (HGC118, HGC111, HGC103, HGC6, HGB4 and HGA8). The colours represent the fluorescent dye chosen for the forward primer: blue – 6FAM, green – VIC, yellow – NED, and red – PET (Applied Biosystems). The bold number adjacent to the bar indicates the size of the repetitive motif.

2.2.4. Statistical Analysis

2.2.4.1. *Descriptive Statistics*

Genetic variation within and between samples was measured by calculating expected (H_e) (Nei 1978) and observed (H_o) heterozygosities using Microsatellite Toolkit (Park 2001). Allelic richness (A_R) was estimated in FSTAT 2.9.3 (Goudet 1995) using the rarefaction method (El Mousadik & Petit 1996). Deviations from Hardy-Weinberg Equilibrium (HWE) were investigated with the Hardy-Weinberg exact test implemented in GENEPOP 4.2.1 using the default settings (Raymond & Rousset 1995; Rousset 2008) and assessed using F_{IS} with relative 95% confidence intervals (CI) calculated by 1000 bootstraps over loci using GENETIX 4.05 (Belkhir et al. 1996-2004). Linkage disequilibrium was examined with the exact test in GENEPOP, using the default settings.

The frequency of null alleles for each locus/sample combination was estimated with FreeNA (Chapuis & Estoup 2007) and the excluding null alleles (ENA) correction was applied to both global and pairwise F_{ST} values (Weir & Cockerham 1984). In order to evaluate the neutrality of the loci used in the genetic analysis, a neutrality test was employed in LOSITAN (Antao et al. 2008), which implements the F_{ST}-outlier method of Beaumont & Nichols (1996). LOSITAN identifies loci that are candidates for balancing and divergent selection and was run under the default settings.

2.2.4.2. *Power Analysis*

In order to assess the statistical power when testing for genetic differentiation, the simulation-based computer program POWSIM (Ryman & Palm 2006) was used, which assesses type I (the probability of rejecting H_O when it is true) and type II (the probability of accepting H_O when it is false) errors. Simulations were run to achieve F_{ST} values ranging from 0 to 0.05, by combining a N_e of 10,000 and different numbers of generations (*t*) (20, 50, 100, 201, 404, 506 and 1026) for a sample size of 48.

2.2.4.3. *Bayesian Clustering Analysis*

Population structuring was investigated using the Bayesian clustering method implemented in STRUCTURE 2.3 (Pritchard et al. 2000; Falush et al. 2003, 2007)

using 200,000 Markov chain Monte Carlo (MCMC) repeats following a burn-in period of 20,000. STRUCTURE groups individuals into genetic clusters by minimising Hardy-Weinberg disequilibrium and linkage disequilibria within groups. Outputs were collated with the web-based program STRUCTURE HARVESTER (Earl & vonHoldt 2012), which enabled the mean likelihoods per K value to be visualised, thus facilitating the detection of the number of genetic clusters that best fit the data.

2.2.4.4. *Genetic Diversity Measures*

Global and pairwise F_{ST} were assessed in FSTAT 2.9.3 (Goudet 1995), where the significance levels, estimated by 1000 permutations, were adjusted for multiple comparisons with a sequential Bonferroni correction (Rice 1989), when required. The more recently proposed genetic diversity indices G'_{ST} (Hedrick 2005) and D_{est} (Jost 2008) were estimated using the web-based application SMOGD (Crawford 2010). G'_{ST} is a standardised measure of genetic differentiation, whereas Jost's D is described as true differentiation (Jost 2008). In both cases, it has been suggested that the indices may poorly reflect population demography where mutation rates are high, as expected for microsatellite markers (Whitlock 2011), however, they are included here for comparison with F_{ST}. Principal Coordinate Analysis (PCoA) was performed in R 3.0.2 (R Core Team 2013) to produce comparative MDS plots of D_{est} and F_{ST}. A Mantel test (Mantel 1967; Smouse et al. 1986) was then implemented, using the R package ECODIST (Goslee & Urban 2007), to measure the correlation between the two genetic distance matrices.

Isolation by distance (IBD) (Wright 1943) was tested using the Isolation by Distance Web Service (IBDWS) (Jensen et al. 2005), which assesses the relationship between the genetic distance matrix (F_{ST}) and a corresponding matrix of geographic distances (km), followed by a Mantel test. Geographic distances were measured by tracing the shortest route between two populations via sea using the GeoDistance website (www.geodistance.com).

2.2.4.5. *Estimating Directional Emigration/Immigration Rates*

Recent migration rates among Lundy (LNTZ), Devon (DEV), south Wales (SW) and southern Ireland (WF & WEX) were calculated using a Bayesian inference approach implemented in the program BAYESASS 3.0.3 (Wilson & Rannala 2003). Simulations with a unique seed value were run for 10,000,000 iterations with MCMC chains sampled every 1,000 iterations, following an initial burn-in of 1,000,000 iterations. As suggested in the program documentation, the five mixing parameters were adjusted to ensure acceptance rates between 20% and 60%, with ΔM, ΔA and ΔF set to 0.70. In order to examine convergence, the posterior mean parameter estimates of multiple runs were compared.

2.2.4.6. *Effective Population Size and Bottleneck Analysis*

The effective population size (N_e) was estimated using LDNE (Waples & Do 2008), in which a random mating model was assumed.

To detect any recent effective population size reductions, BOTTLENECK 1.2 (Piry et al. 1999) was utilised using 10,000 permutations for the Infinite Allele (IAM) (Crow & Kimura 1970), Stepwise Mutation (SMM) (Ohta & Kimura 1973) and Two-Phase Mutation (TPM) (Di Rienzo et al. 1994) models of microsatellite evolution. Approximately 90% of microsatellite mutations are single step (Garza & Williamson 2001), thus the parameters of the TPM model were set to run at 90% of single step mutations with a variance of 10 among multiple steps. The Wilcoxon test was used to determine whether any of the sample sites show an excess of heterozygosity, which is expected after a severe bottleneck (Cornuet & Luikart 1996). In addition, the graphical mode-shift test was incorporated to detect shifts from the normal L-shaped distribution of allele frequencies that are expected at equilibrium (Luikart et al. 1998).

A second approach to assess for evidence of a population bottleneck was the *M*-ratio between the total number of alleles (*k*) and the overall range in allele size (*r*) (Garza & Williamson 2001). When a population has gone through a recent reduction in size, alleles will become lost due to the enhancement of genetic drift. The loss of any allele will cause a reduction in *k*, however, only the loss of the smallest or

largest allele will result in a reduction in *r*. Therefore, *k* is expected to be reduced more quickly than *r* and the resulting *M*-ratio will be smaller in populations that have experienced a recent reduction compared to those at equilibrium (Garza & Williamson 2001). *M* was calculated in ARLEQUIN 3.5 (Excoffier & Lischer 2010) by reporting the mean population value from the modified Garza-Williamson index computation. The modification of the GW index avoids a division by zero when a gene sample is fixed for a single allele (Excoffier et al. 2005). Values of $M < 0.68$ were considered as a sign of a recent population bottleneck (Garza & Williamson 2001), as described in Coscia et al. (2012).

2.2.4.7. *Variation among Groups*

Finally, the variation among groups of samples was tested using FSTAT, with 1000 permutations and a two-sided test, to assess A_R, H_O, gene diversity (H_S), F_{IS}, F_{ST}, relatedness and corrected relatedness. Two different groupings were analysed: (1) males versus females across all populations to test for potential sex differences, and (2) Lundy NTZ versus the other sampled sites to test for site-associated differences.

2.3. RESULTS

2.3.1. Modelling Larval Transport

2.3.1.1. *Hydrodynamic Circulation*

Barotropic forces and residual currents in the Irish and Celtic Seas have been reproduced using the hydrodynamic model to show the effect on larval dispersal (Figure 2.4). There are several residual currents worthy of note, which are: the westward residual current (0.05 – 0.2 m s^{-1}), which flows from South Wales towards Ireland, across St George's Channel, and continuing westward along the south coast of Ireland towards the Celtic Sea; residuals flow east from Lundy Island towards Devon, before interacting with the aforementioned westwards current; and the western Irish Sea gyre, which flows in an anti-cyclonic direction, bounded to the east by North Wales and to the west by Ireland. Residual currents are negligible in Cardigan Bay and in the Celtic Sea.

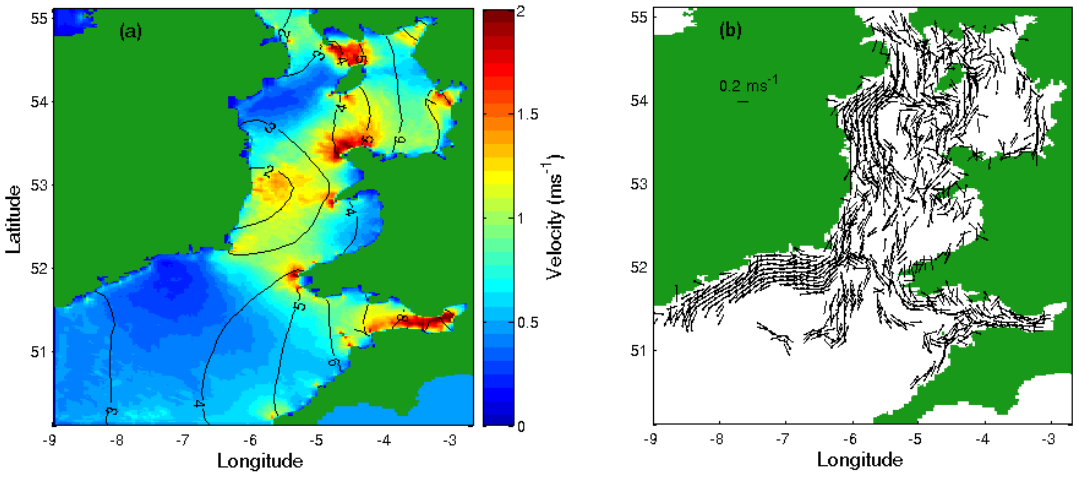


Figure 2.4. (a) Co-tidal contours of maximum tidal range (m) during the 1990 hydrodynamic simulation, superimposed upon coloured contours of maximum depth-averaged scalar tidal velocity (m s^{-1}). (b) Depth-averaged residual currents over the period 1st June-31st September 1990. Residuals less than 0.02 m s^{-1} have been removed for clearer visualisation of the stronger currents. Residual baroclinic currents from the Bristol Channel towards Ireland are depicted. The western Irish Sea gyre (anti-clockwise currents between North Wales and Ireland) is also simulated. Modified from Wootton et al. (in prep).

2.3.1.2. *Forward-tracking Simulations: where do Lundy NTZ larvae go?*

Dispersal variance throughout the release period was small; average dispersal distances only increased by approximately 5% during summer months due to strengthening baroclinic currents. Therefore, all data are calculated from instantaneous larval positions from all release dates.

Dispersal probability distributions and patch-averaged dispersal distances were calculated for each larval behaviour strategy (PTMs 1-6) (Figure 2.5), thus indicating where the larvae are located at the end of each larval stage. Connectivity matrices for larval distributions at the end of each life stage, averaged over all larvae (i.e. 160,000) from each behavioural strategy are shown in Figure 2.6. In addition, connectivity maps have been produced in order to highlight the spatial pattern of connectivity of larvae released from Lundy NTZ in relation to self-recruitment around Lundy Island (Figure 2.7).

The passive larvae simulation (PTM-1), without behavioural influence, provides details on larval dispersal, settlement and connectivity based solely on the physical parameters of the model. In general, a longer PLD led to increased larval dispersal from the NTZ; with the patch-averaged dispersal distances being approximately 15, 21, 27 and 58 km at the end of each life stage, respectively (Figure 2.5a). The larvae were initially transported from Lundy towards the Devon coast, by eastward residual currents. Subsequently, the larvae were advected south along the English coast, and also northwards across the Bristol Channel towards south Wales. The connectivity matrix, which details settlement potential reveals that between 22% and 34%, depending on PLD, of passive larvae were able to settle, with the majority along Devon and Cornwall coasts (Figure 2.6a). Early settlers (stages I and II) were more likely to settle at Devon, whilst late settlers (stages III and IV) would settle on the Cornwall coast. No passive larvae settled on Welsh or Irish coastlines (Figure 2.7a). Potential for self-recruitment at Lundy was very low (< 2%) and the longer the PLD, the reduced likelihood of settlement. However, slightly more larvae settled at Lundy West (RZ) than at the Lundy East (NTZ).

Larval behavioural strategies (PTMs 2-6) reveal a strong influence on dispersal, with each strategy producing different dispersal patterns to the passive strategy (PTM-1).

Larvae positioned in the upper water column (PTMs 2-3) showed similar patterns, as larvae were initially transported eastwards towards the Devon coast during stages I and II, yet during stages III and IV larvae migrated westwards, due to entrainment in the Celtic Sea front residual current, towards southern Ireland (Figure 2.5b, Figure 2.5c). As such, the majority of settlement occurred along the coasts of Devon in the early stages but by the end of stage IV settlement was only achievable in southern Ireland (Figure 2.7b, Figure 2.7c). There was also a small settlement along south west Wales during stages II and III (Figure 2.6b, Figure 2.6c). Patch-averaged dispersals revealed long distance dispersal, with approximately 185 km being attained by the end of stage IV. Larvae positioned at 2.5 m (PTM-2) dispersed approximately 10% further than larvae at 5 m (PTM-3). In terms of self-recruitment, settlement around Lundy was low (< 1.6 %).

Larvae located at mid-depth in the water column (PTMs 4-5) were advected along different pathways to those located in the upper water column (PTMs 2-3), attributable to the residual currents at mid-depths in the outer Bristol Channel and around South Wales being in a more northerly direction than those at the surface. The larvae in PTM-4 and PTM-5 generally travelled eastwards to Devon (Figure 2.5d, Figure 2.5e) and only moved north towards Wales during stages III and IV because of entrainment in northwards mid-depth residual currents. Consequently, connectivity matrices show high settlement along Devon coasts during all life stages, with PTM-5 (31-57%) achieving more than a two-fold increase in settlement compared with PTM-4 (15-18%) (Figure 2.6d, Figure 2.6e). Similar to the upper water column larvae (PTMs 2-3), there was a small settlement along south west Wales, although none of the mid-depth larvae travelled west towards Ireland. At all life stages, relatively high proportions of larvae (< 5.33%) settled around Lundy, particularly in the NTZ (Figure 2.7d, Figure 2.7e). In fact, PTM-4 shows the highest settlement at Lundy (i.e. self-recruitment) out of all investigated behavioural strategies (Figure 2.6d).

The diel behavioural strategy (PTM-6) reveals a different dispersal pathway to other PTMs (Figure 2.5f). Initially, larvae were transported eastwards by residual currents towards Devon, then during stage III they began to migrate west towards Ireland.

Connectivity data reveals high settlement (> 20%) in Devon during early stages (Figure 2.6f). However, this was dramatically reduced during stages III and IV. There was a small settlement (3.14%) in southern Ireland by the end of stage IV, but overall, results suggest that if larvae did not settle in Devon during stages I-III, they are likely to become stranded off-shore in deep water, thus being unsuccessful. Dispersal distances during stages I-III were short (< 50 km), however, by the end of stage IV, larvae had travelled approximately 100 km from Lundy NTZ, as they became entrained in the Celtic Sea front residual current (Figure 2.7f). Self-recruitment was low, with maximum settlement being reached during stage II (Figure 2.6f).

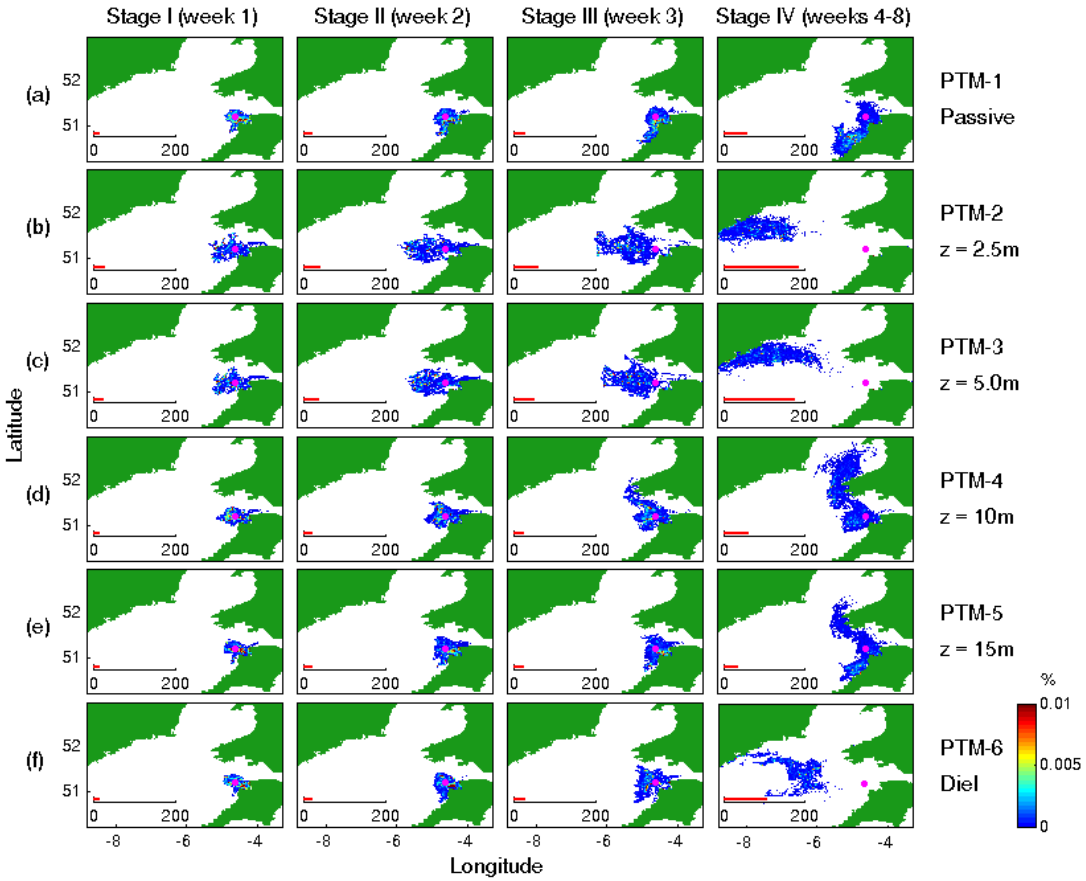


Figure 2.5. *Present-day scenario: 1990*. Dispersal probability distributions (dispersal kernels) for lobster larvae released from Lundy Island NTZ (east coast, denoted by magenta circles). Each panel shows probability distributions for all simulations over the spawning season (May - September), totalling 160,000 separate particle trajectories. PTM simulations lasted for a PLD of 56 d (8 weeks) and probability distributions after Stage I (7 d), Stage II (14 d), Stage III (21 d) and Stage IV (56 d) are presented (columns). Each row represents a different larval behaviour scenario (PTM 1: passive larvae, PTM 2: z ≈ 2.5 m, PTM 3: z ≈ 5 m, PTM 4: z ≈ 10 m, PTM 5: z ≈ 15 m, PTM 6: diel migration), where z indicates the depth larvae were positioned for Stages I-III. In all cases, larvae were in the surface layer for Stage IV. The colour scale indicates the proportion of total particles located in each model cell (white cells indicate no larvae present and green cells represents land). The bar graphs in each figure indicate the patch averaged dispersal distance from the NTZ, at the end of each Stage. Modified from Wootton et al. (in prep).

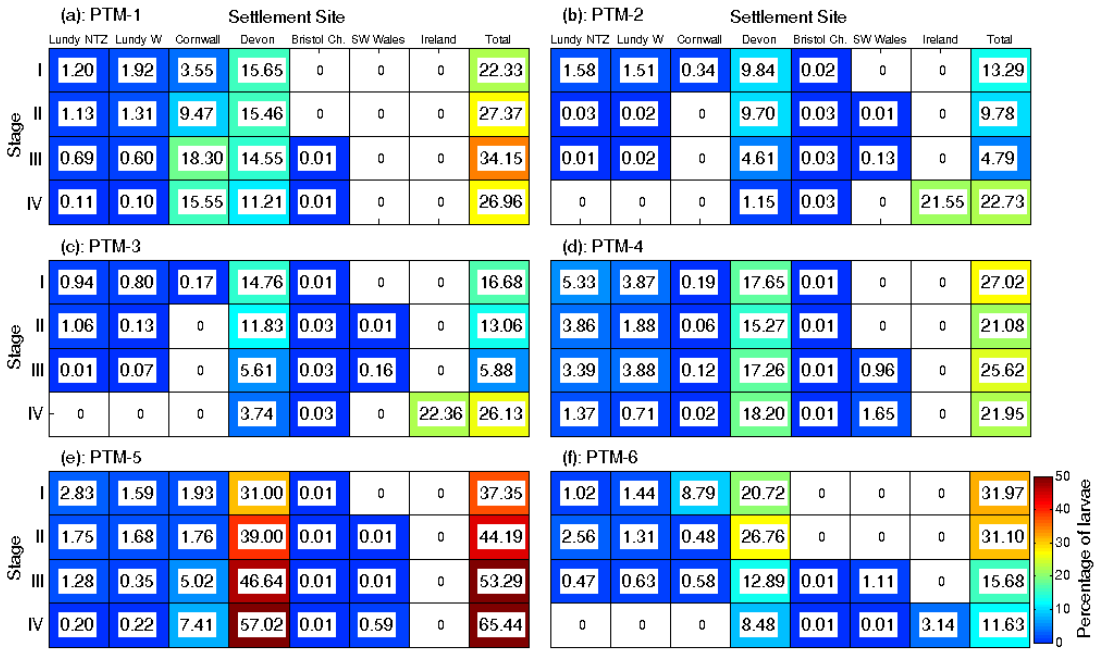


Figure 2.6. *Present-day scenario: 1990*. Connectivity matrices for lobster larvae released from Lundy Island NTZ. (a) PTM-1 (passive), (b) PTM-2 (z ≈ 2.5 m), (c) PTM-3 (z ≈ 5 m), (d) PTM-4 (z ≈ 10 m), (e) PTM-5 (z ≈ 15 m), and (f) PTM-6 (diel migration). Each matrix shows percentages of all larvae (averaged over 16 releases) that settle on the 5 coastlines (columns). Each row represents the end of the specified larval stage. The total proportions of settled particles are indicated in the right-hand columns of each matrix. Modified from Wootton et al. (in prep).

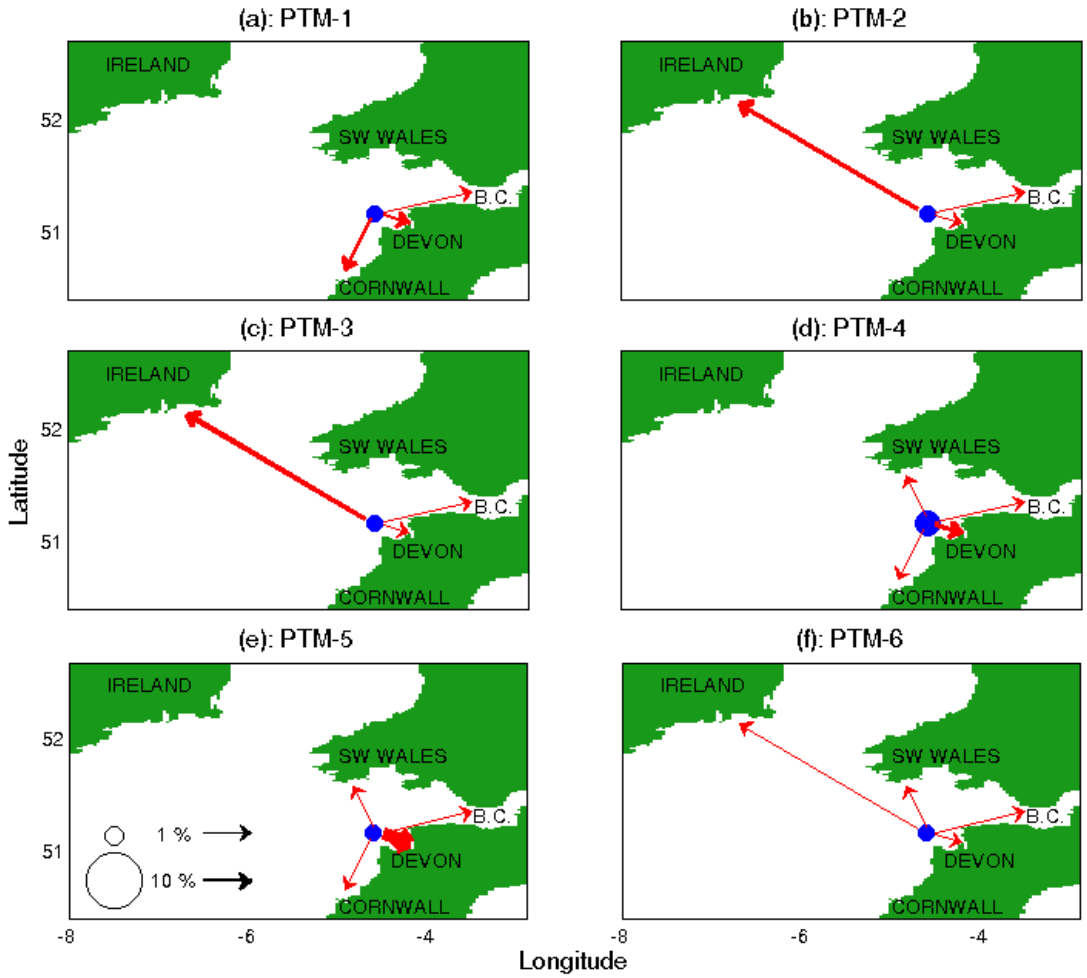


Figure 2.7. *Present-day scenario: 1990*. Connectivity maps showing levels of self-recruitment (circles) at Lundy Island (combined Lundy NTZ and Lundy West), and connectivity (arrows) between Lundy NTZ and other coastal populations, at the end of each PTM simulation (i.e. after a PLD of 8 weeks). (a) PTM 1: passive larvae, (b) PTM 2: z ≈ 2.5 m, (c) PTM 3: z ≈ 5 m, (d) PTM 4: z ≈ 10 m, (e) PTM 5: z ≈ 15 m, (f) PTM 6: diel migration, where z indicates the depth larvae were positioned for Stages I-III. In all cases, larvae were in the surface layer for Stage IV. Each figure shows connectivity/recruitment averaged over all simulations throughout the spawning season (May - September), totalling 160,000 separate particle trajectories. Modified from Wootton et al. (in prep).

2.3.1.3. *Backwards-tracking Simulations: where do Lundy NTZ settlers come from?*

Very few simulated passive larvae that settle at Lundy NTZ (< 5%) originated from coastal sites, although the patch did remain close to Lundy (< 25 km away) during early life stages (stages I-III) (Figure 2.8a). This suggests that passive larvae that settle at Lundy, within three weeks, are likely to have been released around Lundy or off-shore from Devon. Passive larvae that settle at Lundy at the end of stage IV (8 weeks) were most likely to have been released off-shore to the west and within 50 km of Lundy.

Lobster larvae that employ one of the modelled behaviour strategies (PTMs 2-6) were most likely to have originated from the east, specifically the Devon coast and elsewhere within the Severn estuary (Figure 2.8b-f). For PTM-2 and PTM-3, more than 50% of the larvae were likely to have been released from North Devon, and significant proportions (~ 10%) released from the Bristol Channel during stage IV. Maximum dispersal distances of larvae near the surface were approximately 50 km. Comparing these distances with larvae released at Lundy (~ 180 km at the end of stage IV) illustrates the strong influence of the Celtic Sea front residual current on larval dispersal. Larvae positioned in mid-depth waters (PTM-4 and PTM-5) were again likely to have been released from North Devon. However, a higher proportion of larvae are self-recruiting around Lundy Island (0.5 – 3.5 % self-recruitment), compared with less than 0.5 % for larvae positioned at the surface (results not shown). Larvae adopting these mid-depth strategies were also likely to have been released off-shore to the west and within approximately 100 km of Lundy. Diel migration (PTM-6) showed a similar source distribution and dispersal distances to the other strategies, although significant proportions of larvae could also migrate from southwest Wales to Lundy during stages II-IV.

For all simulations, it was not possible for larvae from further afield than 100 km (e.g. North Wales, Ireland and France) to reach Lundy Island within the modelled 56 days PLD.

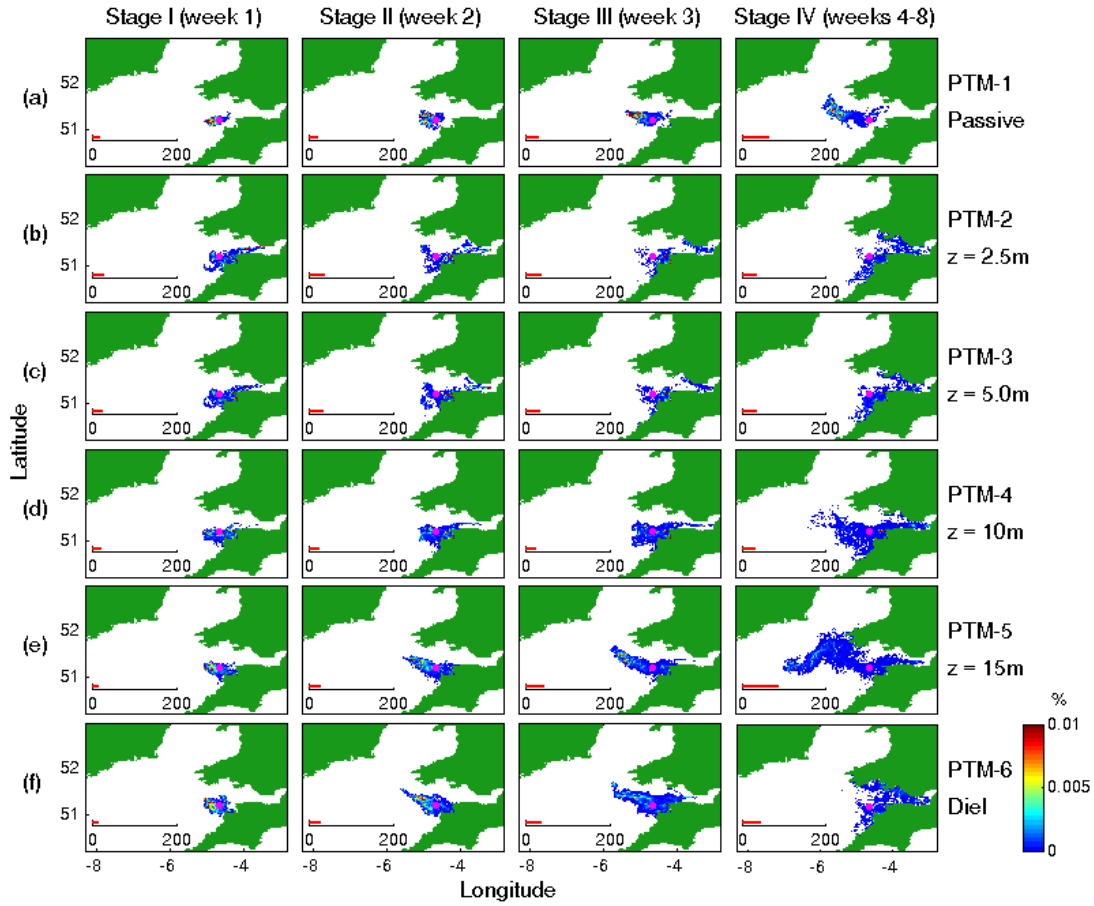


Figure 2.8. *Present-day (particle backtracking) scenario: 1990*. Source-probability distributions (calculated from particle backtracking simulations) of lobster larvae that settle at the Lundy NTZ (denoted by magenta circles). Each panel shows source-probability distributions for all simulations (May – September 1990), totalling 160,000 separate particle trajectories. Particle backtracking simulations were run for a PLD of 8 weeks, and source-probability distributions after Stage I (7 d), Stage II (14 d), Stage III (21 d) and Stage IV (56 d) are presented (columns) (each larval stage was run backwards, separately). Each row represents a different behavioural scenario (PTM 1: passive larvae, PTM 2: z ≈ 2.5 m, PTM 3: z ≈ 5 m, PTM 4: z ≈ 10 m, PTM 5: z ≈ 15 m, PTM 6: diel migration), where z indicates the depth larvae were at for Stages I-III. In all cases, larvae were in the surface layer for Stage IV. The colour scale indicates the proportion of total particles located in each model cell (white cells indicate no larvae present and green cells represents land). The bar graphs in each figure indicate the patch averaged dispersal distance from the NTZ, at the end of each Stage. Modified from Wootton et al. (in prep).

2.3.2. Genetic Diversity and Population Structure

H_e ranged from 0.668 in the WEX population to 0.690 in the DEV population and H_o ranged from 0.639 in the LNTZ population to 0.694 in the SW population. NW was found to have the highest N_A across loci with 8.00, whereas CB had the lowest with 7.00. Also, LICZ had the highest A_R with 7.42, while NW had the lowest with 6.97 (Table 2.3). Although, locus HGA8 exhibited deviations from HWE at 5 out of the 9 populations tested ($P < 0.05$), indicating potential technical artefacts such as null alleles. For the remaining loci, no locus showed deviations from HWE in more than 2 populations (Table 2.4). No populations exhibited significant deviations from HWE (Table 2.3).

Table 2.3. Genetic diversity parameters inferred from microsatellite markers for *Homarus gammarus* in the Irish Sea. H_e and H_o expected and observed heterozygosities, N_A number of alleles, A_R allelic richness, and F_{IS} inbreeding coefficient.

	H_e	H_o	N_A	A_R	F_{IS}
LNTZ	0.688	0.639	7.75	7.17	0.072
LICZ	0.681	0.650	8.00	7.42	0.046
WF	0.675	0.642	7.92	7.24	0.050
CB	0.670	0.663	7.58	7.07	0.012
ND	0.687	0.676	7.25	6.99	0.016
WEX	0.668	0.644	7.00	7.00	0.037
SW	0.683	0.694	7.83	7.15	-0.017
NW	0.674	0.642	7.92	7.30	0.049
DEV	0.690	0.681	7.50	6.97	0.014

Table 2.4. Hardy-Weinberg probability test for *Homarus gammarus* in the Irish Sea. Significant values are marked with an asterisk (P < 0.05).

PHWE	HGD111	HGD106	HGC131b	HGC129	HGC120	HGB6
LNTZ	0.314	0.057	0.440	0.177	0.180	0.940
LICZ	0.117	0.630	0.270	**0.005***	0.380	0.136
WF	0.445	0.313	0.497	0.193	0.662	0.279
CB	0.571	0.896	0.482	0.892	0.195	0.273
ND	0.418	0.091	0.569	0.552	0.132	0.620
WEX	0.737	0.386	0.320	0.940	**0.009***	0.948
SW	0.288	0.758	0.221	0.798	0.072	0.678
NW	0.378	0.344	0.819	0.321	0.156	0.706
DEV	0.759	0.504	0.619	0.897	**0.007***	0.753

PHWE	HGC118	HGC111	HGC103	HGC6	HGB4	HGA8
LNTZ	0.070	0.701	0.340	0.113	0.338	**0.000***
LICZ	0.099	0.375	0.983	0.353	0.641	**0.000***
WF	0.264	0.292	0.999	0.798	0.088	**0.006***
CB	0.785	0.375	0.322	**0.006***	0.245	0.097
ND	0.380	0.113	0.478	1.000	0.084	**0.002***
WEX	0.742	0.697	0.571	0.314	0.239	**0.043***
SW	0.539	0.722	0.571	1.000	0.486	0.063
NW	0.239	0.870	0.747	1.000	0.978	0.152
DEV	0.608	0.687	0.650	0.478	0.417	0.780

Three locus-pairings were found to be in global linkage disequilibrium: HGD106 & HGC131b (P = 0.026); HGC129 & HGC118 (P = 0.022); and HGD111 & HGC103 (P = 0.040) (Appendix 1). However, for both the HGD106 & HGC131b and HGD111 & HGC103 locus-pairings, the significant global test results were seemingly due to significant results in a small number of samples (and became non-significant when these samples were omitted). Only one population was significant in the HGD106 & HGC131b locus-pairing (WEX, P = 0.003), while only two populations were significant in the HGD111 & HGC103 locus-pairing (SW, P = 0.002; NW, P = 0.012). There were no significant single sample linkage results for the HGC129 & HGC118 locus-pairing, which can be reported as a Type I error. Subsequent pairwise analyses were repeated omitting one locus and results were not qualitatively different.

Locus HGA8 was highlighted as having a large chance ($r \geq 0.20$) of harbouring null alleles at LNTZ with r = 0.21. F_{IS} values were recalculated and tested for significance

with and without HGA8 and there was no change in the level of significance at any population, with all populations remaining non-significant. The remaining populations and all other loci were shown to have a moderate ($0.05 \leq r < 0.20$) to negligible ($r < 0.05$) chance of null alleles (Table 2.5). When the ENA correction was applied to the global and pairwise F_{ST} values, all comparisons remained non-significant (Appendix 2). Global F_{ST} before ENA correction was -0.000305 (95% CI: -0.001725 - 0.001216) and was 0.000105 (95% CI: -0.001345 - 0.001659) after ENA correction, with a value of -0.000 (95% CI: -0.002 – 0.001) when calculated in FSTAT. Correspondingly, the pairwise F_{ST} values calculated in FSTAT were non-significant after Bonferroni correction (Table 2.6).

Table 2.5. The estimate of null allele frequency per locus per population of *Homarus gammarus* in the Irish Sea using FreeNA (Bootstrap = 1000). The values in white, light grey and dark grey squares represent a negligible ($r < 0.05$), moderate ($0.05 \leq r < 0.20$) or large ($r \geq 0.20$) chance of containing null alleles, respectively.

	Sampling Locations								
Loci	**LNTZ**	**LICZ**	**WF**	**CB**	**ND**	**WEX**	**SW**	**NW**	**DEV**
HGD111	0.010	0.020	0.006	0.000	0.000	0.000	0.000	0.000	0.000
HGD106	0.062	0.000	0.016	0.000	0.000	0.007	0.000	0.000	0.000
HGC131b	0.000	0.013	0.000	0.000	0.000	0.055	0.021	0.022	0.000
HGC129	0.047	0.080	0.010	0.027	0.000	0.000	0.000	0.064	0.011
HGC120	0.018	0.042	0.059	0.033	0.000	0.066	0.074	0.000	0.046
HGB6	0.000	0.042	0.000	0.025	0.000	0.000	0.000	0.000	0.000
HGC118	0.064	0.000	0.009	0.000	0.000	0.000	0.000	0.000	0.000
HGC111	0.024	0.000	0.031	0.000	0.001	0.000	0.000	0.031	0.000
HGC103	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.019
HGC6	0.017	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HGB4	0.033	0.028	0.003	0.000	0.040	0.026	0.033	0.000	0.028
HGA8	0.210	0.134	0.108	0.053	0.099	0.120	0.009	0.073	0.028

Table 2.6. F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Homarus gammarus* in the Irish Sea.

	LNTZ	**LICZ**	**WF**	**CB**	**ND**	**WEX**	**SW**	**NW**	**DEV**
LNTZ		0.669	0.364	0.285	0.293	0.668	0.106	0.511	0.143
LICZ	-0.002		0.563	0.285	0.547	0.756	0.699	0.942	0.715
WF	0.002	-0.002		0.447	0.165	0.410	0.458	0.317	0.711
CB	0.000	0.001	0.002		0.176	0.658	0.486	0.893	0.682
ND	-0.002	-0.001	0.004	-0.001		0.568	0.586	0.899	0.371
WEX	-0.001	0.001	0.004	0.000	0.003		0.569	0.818	0.414
SW	0.001	-0.002	0.000	-0.001	0.000	0.001		0.972	0.857
NW	0.000	-0.002	0.004	-0.004	-0.003	0.001	-0.002		0.956
DEV	0.000	-0.001	0.000	-0.001	-0.002	0.003	-0.003	-0.004	

LOSITAN confirmed the neutrality of the 12 microsatellite markers that were used, with none being selected as candidates for balancing or positive selection (Figure 2.9). POWSIM analysis indicated that the data (loci and sample sizes) conferred a low Type I error (P = 0.04) and a high probability (P = 0.99) for detecting differentiation at F_{ST} = 0.01 (Table 2.7).

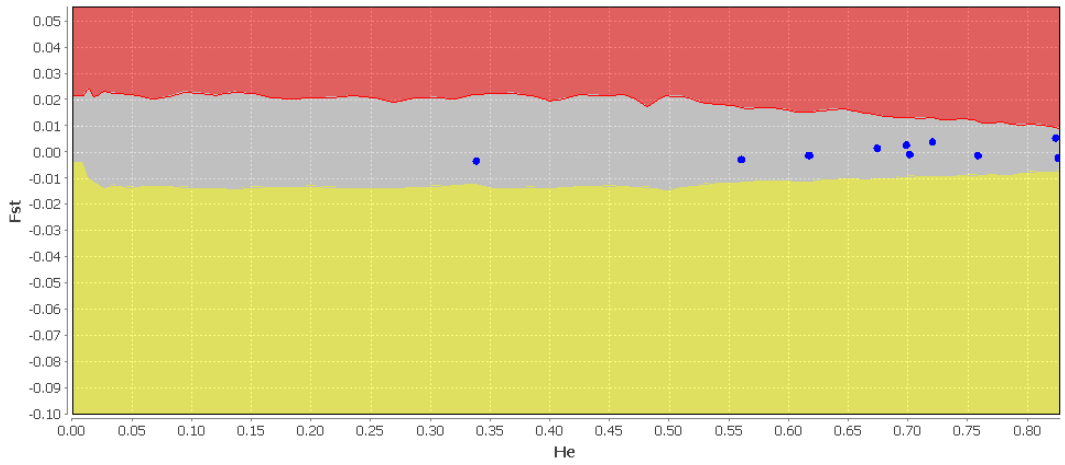


Figure 2.9. LOSITAN neutrality test for the 12 microsatellite makers (blue dots) used on *Homarus gammarus* in the Irish Sea. Markers are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area).

Table 2.7. Results of POWSIM analysis indicating the probability of detecting differentiation by exact tests at various levels of simulated true structuring (F_{ST} 0.001-0.05). Proportion of significant tests at F_{ST} = 0 indicate the Type I error probability.

F_{ST}							
0	**0.001**	**0.0025**	**0.005**	**0.01**	**0.02**	**0.025**	**0.05**
0.04	0.15	0.34	0.70	0.99	1	1	1

Overall, there was no evidence of any population structuring, as there was a mean likelihood of K = 1 (LnP(K) = -13617.18) (Figure 2.10) and an equal assignment of each individual to K clusters (Figure 2.11).

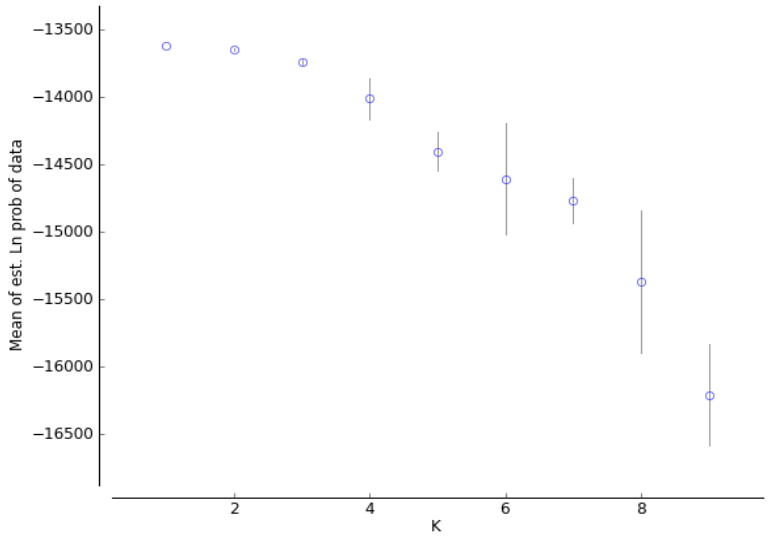


Figure 2.10. Plot of mean likelihood L(K) and variance per K value from STRUCTURE for *Homarus gammarus* in the Irish Sea containing 385 individuals genotyped for 12 microsatellite loci.

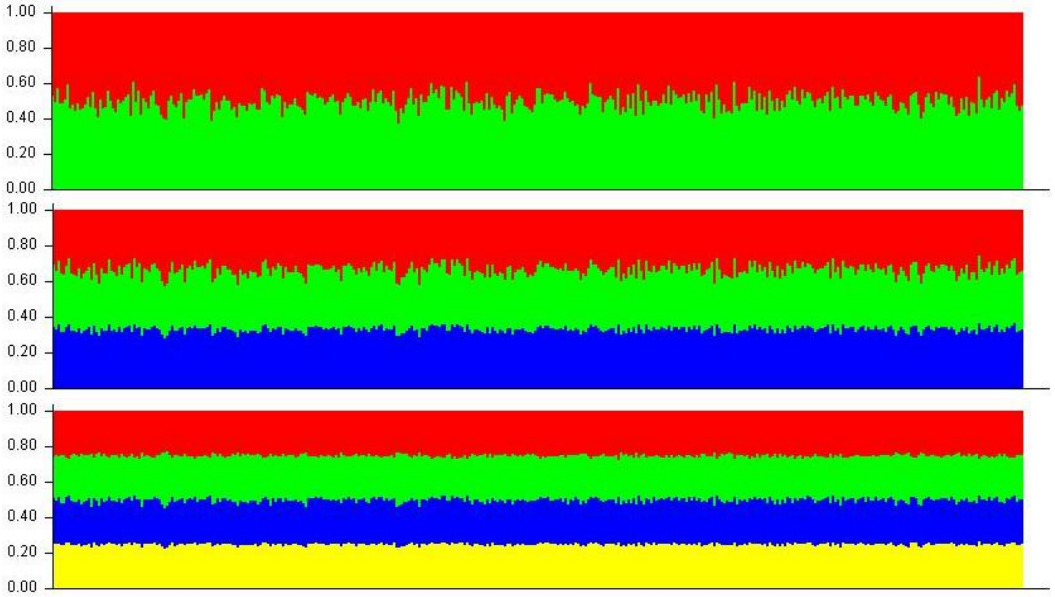


Figure 2.11. STRUCTURE bar plots for *Homarus gammarus* in the Irish Sea for K = 2, K = 3 and K = 4 for 385 individuals showing assignment to K clusters. Each vertical bar represents an individual and each colour a different cluster.

Locus-specific G'_{ST} and D_{est} values were very low, with HGA8 shown to have the most genetic differentiation in both measures but with < 2.9% of the maximum possible (G'_{ST} = 0.029, D_{est} = 0.025). Eight loci revealed no genetic differentiation (zero values) in either measure, plus HGD106 (G'_{ST} = 0.005, D_{est} = 0.003), HGC111 (G'_{ST} = 0.013, D_{est} = 0.010) and HGC103 (G'_{ST} = 0.009, D_{est} = 0.006) were only slightly differentiated (Table 2.8). Population pairwise D_{est} revealed low levels of differentiation that were similar to those shown by F_{ST} (r = 0.861, P < 0.001) (Table 2.9). The MDS plots for the two genetic diversity indices were visually similar, as LNTZ, LICZ and SW were clustered together in both, with the other populations appearing scattered (F_{ST} – Figure 2.12, D_{est} – Figure 2.13).

Table 2.8. Locus-specific G'_{ST} and D_{est} values for *Homarus gammarus* in the Irish Sea.

	G'_{ST}	D_{est}
HGD111	0.000	0.000
HGD106	0.005	0.003
HGC131b	0.000	0.000
HGC129	0.000	0.000
HGC120	0.000	0.000
HGB6	0.000	0.000
HGC118	0.000	0.000
HGC111	0.013	0.010
HGC103	0.009	0.006
HGC6	0.000	0.000
HGB4	0.000	0.000
HGA8	0.029	0.025

Table 2.9. D_{est} pairwise comparisons for *Homarus gammarus* in the Irish Sea.

D_{est}	LNTZ	LICZ	WF	CB	ND	WEX	SW	NW	DEV
LNTZ									
LICZ	-0.00007								
WF	0.00078	0.00004							
CB	0.00000	0.00034	0.00090						
ND	-0.00016	0.00000	0.00319	-0.00004					
WEX	-0.00007	0.00000	0.00216	0.00000	0.00004				
SW	0.00002	-0.00123	0.00002	-0.00003	-0.00002	0.00000			
NW	0.00000	-0.00001	0.00155	-0.00408	-0.00163	0.00001	-0.00021		
DEV	0.00000	0.00000	0.00001	-0.00008	-0.00008	0.00061	-0.00183	-0.00274	

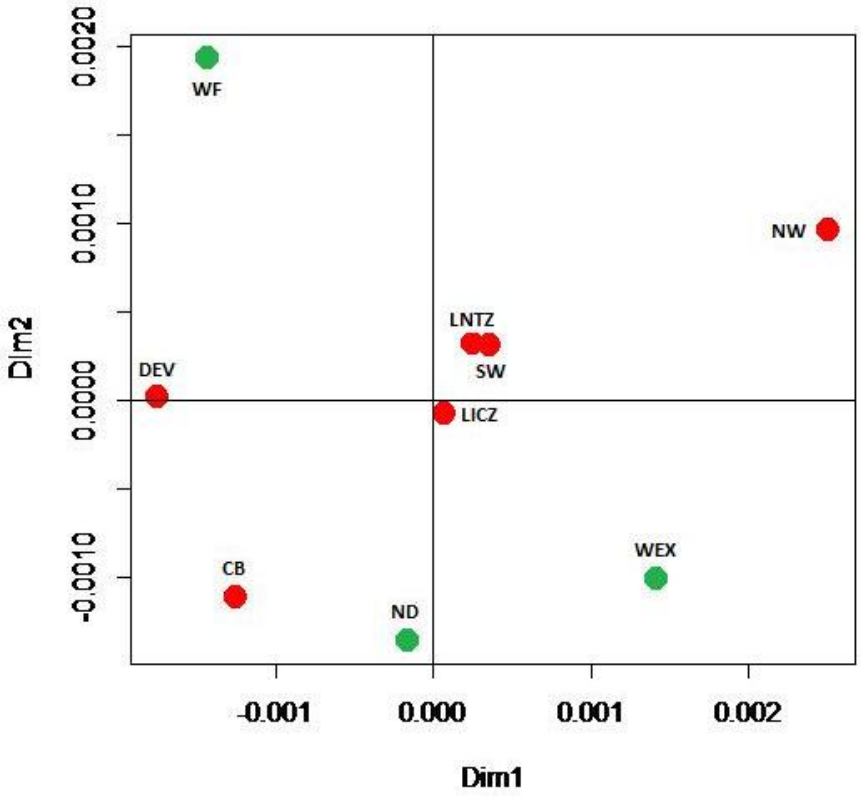


Figure 2.12. MDS plot of pairwise F_{st} values of *Homarus gammarus* in the Irish Sea. Red circles are eastern Irish Sea localities and green denotes western Irish Sea localities.

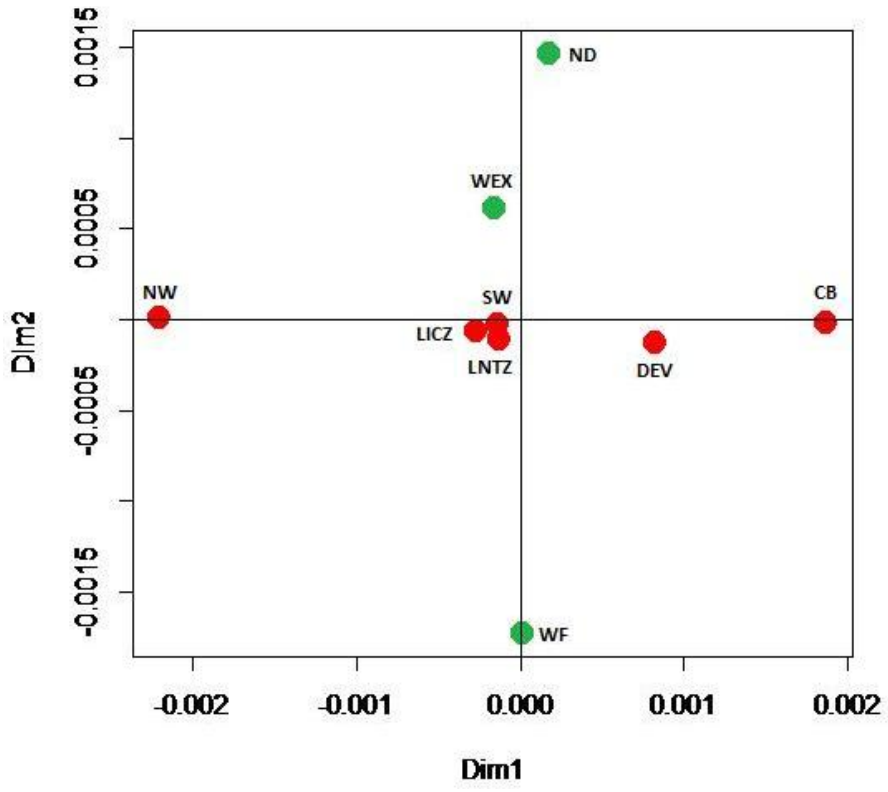


Figure 2.13. MDS plot of pairwise D_{est} values of *Homarus gammarus* in the Irish Sea. Red circles are eastern Irish Sea localities and green denotes western Irish Sea localities.

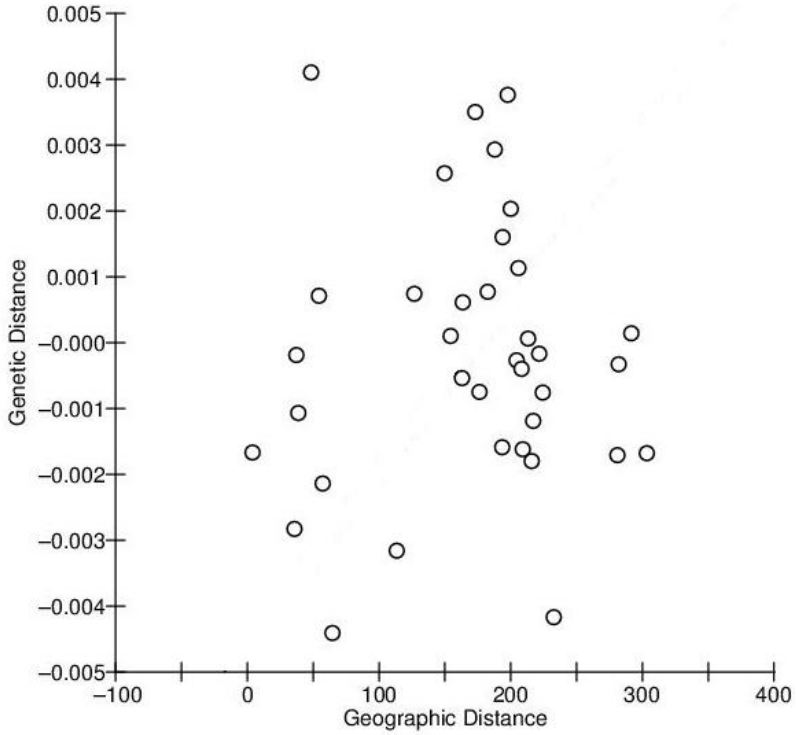


Figure 2.14. IBD in *Homarus gammarus* in the Irish Sea, with genetic distance (F_{ST}) and geographic distance (km).

There was no evidence of IBD (r = 0.057, P = 0.35) (Figure 2.14, Appendix 3).

Inconsistencies between the posterior mean parameter estimates for the BAYEASS simulations indicated there was insufficient information in the data to accurately infer recent migration rates among LNTZ, SW, DEV, and WF & WEX (Appendix 4).

Estimates of effective population size suggest that the sampled populations were very large, with upper confidence intervals including infinity (Table 2.10). The population with the lowest N_e was WEX with 677.7 (95% CI: 98.3 - ∞). BOTTLENECK detected SW (P = 0.032) and DEV (P = 0.017) as populations with potential excess heterozygosity, using the IAM model (Table 2.11). However, the allele frequency distribution was the normal L-shape, which is expected under mutation-drift equilibrium and not the result of a recent bottleneck. The *M*-ratio values were below the threshold for all populations, signifying that populations may have undergone a recent population reduction (M_{LNTZ} = 0.20, M_{LICZ} = 0.21, M_{WF} = 0.21, M_{CB} = 0.20, M_{ND} = 0.19, M_{WEX} = 0.18, M_{SW} = 0.20, M_{NW} = 0.21, M_{DEV} = 0.20).

Table 2.10. Estimates of effective population size (N_e) in *Homarus gammarus* in the Irish Sea, with 95% confidence intervals (parametric). The lowest allele frequency used was 0.02.

	N_e	95% CI
LNTZ	-406.8	642.2 - ∞
LICZ	1252.1	175.6 - ∞
WF	-768.4	400.7 - ∞
CB	-1234.9	253.3 - ∞
ND	-272.7	474.0 - ∞
WEX	677.7	98.3 - ∞
SW	-924.1	378.0 - ∞
NW	2407.4	187.2 - ∞
DEV	-916.0	363.0 - ∞
TOTAL	-1775.9	-7162.0 - ∞

Table 2.11. Assessing heterozygosity excess in *Homarus gammarus* in the Irish Sea using a Wilcoxon test under the Infinite Allele Model (IAM), the Stepwise Mutation Model (SMM) and the Two-Phase Mutation Model (TPM) (90%, variance 10). Significant values are marked with an asterisk ($P < 0.05$).

	IAM	SMM	TPM
LNTZ	0.117	1.000	0.993
LICZ	0.076	1.000	0.999
WF	0.151	1.000	0.998
CB	0.102	1.000	1.000
ND	0.055	0.999	0.995
WEX	0.055	1.000	0.998
SW	**0.032***	1.000	0.998
NW	0.133	1.000	0.999
DEV	**0.017***	1.000	0.997

There was no evidence for sex differences between male and female *H. gammarus* in the Irish Sea in any of the tested diversity parameters (Table 2.12). However, when comparing Lundy NTZ against all other sites F_{IS} (p = 0.013) and Relc (p = 0.010) were revealed as significant (Table 2.13).

Table 2.12. Testing for sex differences between male and female *H. gammarus* across all sampling locations in the Irish Sea. A_R allelic richness, H_o observed heterozygosity, H_S gene diversity, F_{IS} inbreeding coefficient, F_{ST}, Rel relatedness and Relc corrected relatedness.

	A_R	H_O	H_S	F_{IS}	F_{ST}	Rel	Relc
Males	5.240	0.652	0.676	0.036	0.003	0.007	-0.075
Females	5.336	0.667	0.682	0.021	0.001	0.001	-0.043
P-value	0.263	0.243	0.266	0.442	0.256	0.261	0.439

Table 2.13. Comparing Lundy NTZ against all other sites within the Irish Sea. A_R allelic richness, H_o observed heterozygosity, H_S gene diversity, F_{IS} inbreeding coefficient, F_{ST}, Rel relatedness and Relc corrected relatedness.

	A_R	H_O	H_S	F_{IS}	F_{ST}	Rel	Relc
Lundy NTZ	6.724	0.639	0.689	0.072	-0.000	-0.000	-0.154
Other Populations	6.618	0.663	0.679	0.025	-0.000	-0.001	-0.051
P-value	0.276	0.232	0.175	0.013	0.989	0.989	0.010

2.4. DISCUSSION

In this study, population genetic structuring revealed by 12 microsatellite loci was used to describe realised gene flow patterns and compared with spatial patterns of larval recruitment inferred through biophysical modelling for *H. gammarus* in the Irish Sea. This comparison was used as a preliminary validation of the larval biophysical model itself, as well as to provide insight into spatial recruitment dynamics within the region. Particular emphasis was placed on recruitment patterns associated with the Lundy NTZ and its potential role as a source/sink.

The genetic data support the hypothesis of a single panmictic *H. gammarus* population, at least on ecological timescales, within the Irish and Celtic Seas. STRUCTURE analysis without *a priori* sample partitioning provided no evidence of more than one genetic cluster. While the resolution of such clustering analyses may be limited where underlying genetic structure is low (Latch et al. 2006), this was also supported by global and pairwise tests (F_{ST}, exact) among samples. Simulation analysis indicated that such analyses had considerable power to detect even low levels of differentiation. For many marine species, estimates of genetic structure derived from sampling may be compromised by adult dispersal. However, as *H. gammarus* adults are largely sedentary (Jensen et al. 1993; Bannister et al. 1994; Smith et al. 2001; Moland et al. 2011; Øresland & Ulmestrand 2013) this is unlikely to obscure results. Furthermore, genetic homogeneity was consistent across analyses for both sexes and no IBD was detected.

The lack of genetic structuring in the Irish and Celtic Seas is consistent with existing data for the species describing low or absent genetic structuring among *H. gammarus* in the same area (Prodöhl et al. 2007). Yet, the present study adds to previous ones, since the Irish Sea has not formerly been an area of focus. For instance, Ferguson (2002) surveyed very few locations, and none from Lundy Island or the Bristol Channel, within the Irish Sea, and with only six microsatellite loci. While wider scale studies have similarly reported low levels of genetic structure (Ferguson 2002; Jørstad et al. 2005; Triantafyllidis et al. 2005), the chaotic patterns resolved have been interpreted as compatible with non-migration-drift equilibrium, with discrete populations retaining historical patterns of co-ancestry due to

expansion from a single refugium. The M-ratio values generated in the present study may be compatible with Pleistocene induced population reductions, with values lower than the threshold for all populations. However, the critical M value was assumed and not calculated specifically, which needs to be taken into consideration, as it might hinder the final interpretation of this test. In contrast, the Gower (SW) and Ilfracombe (DEV) were the only sampling locations to show the signs of a possible recent population bottleneck. However, this was not considered as strong evidence for a bottleneck, as test results were only significant under the IAM. This mutation model is deemed unlikely for microsatellites (Piry et al. 1999) and it is recommended to use the SMM to be statistically conservative when testing for recent bottlenecks (Luikart & Cornuet 1998).

The prevalence of non-equilibrium conditions has important implications for the results here, as it could mean that contemporary gene flow is overestimated by neutral genetic markers (Marko 2004). Furthermore, particularly for large populations such as those expected for *H. gammarus*, low migration rates might be sufficient to homogenise genetic variation between populations but fail to prevent demographic isolation of populations on time scales of interest to management (Hauser & Carvalho 2008). To address this, biophysical modelling was also employed to assess contemporary dispersal patterns under a range of scenarios. In general, the model predicts connectivity between *H. gammarus* populations at Lundy NTZ and Devon during early development (stages I and II), regardless of the PTM strategy, with large differences in dispersal only becoming apparent in the later stages (III and IV); dispersal probability distributions were along the coasts of Devon, Wales or Ireland by stage IV. The varied dispersal distributions confirm the potentially strong influence of larval behaviour on dispersal, but the limited knowledge of *H. gammarus* larval behaviour and ecology means that it is not possible to select the PTM strategy that most accurately reflects the dispersal of this species. The results from the microsatellite analysis should assist with choosing which strategy is the one most likely adopted by *H. gammarus*, however, the low differentiation in *H. gammarus* throughout the entire sampled range is not reflected by a specific PTM. Moreover, the backwards-tracking simulations implied

that it was not possible for larvae from further afield than 100 km to reach Lundy Island within the modelled 56 days PLD, yet there is low genetic differentiation between Lundy and populations in North Wales and Ireland (Howth, North Dublin) (> 100 km away). This therefore indicates that rather than *H. gammarus* being a panmictic population within the Irish Sea, it is more likely to conform to a type of stepping stone model of population connectivity but that the detection of such patterns is compromised by levels of gene flow that are sufficiently high to homogenise genetic variation within the region and obscure finer scale patterns.

It can be seen from the larval dispersal model that the predominant controlling factors on dispersal were the persistent residual currents, the positioning of the larvae within the water column, and the PLD. Therefore, it is clear that both hydrodynamic and behavioural parameters play fundamental roles in larval dispersal. The passive particle simulation allowed physical controls on larval dispersal to be evaluated, and residual circulation patterns were found to play a fundamental role. However, the results highlight that knowledge of larval behaviour is crucial in predicting larval dispersal, revealing that biological controls determine larval trajectory, and overall dispersal, to a greater extent than physical controls of the water body. This has been shown elsewhere and is in fact a commonly supported theory in planktonic larvae dispersal (Kingsford et al. 2002; Largier 2003; Metaxas & Saunders 2009). For instance, the incorporation of behaviour into biophysical modelling of the American lobster, *H. americanus*, resulted in the successful determination of larval sources and sinks, along with the conclusion that, in addition to larval behaviour, time of spawning, water temperature and prevailing winds were key determinants in larval dispersal (Katz et al. 1994; Incze & Naimie 2000; Harding et al. 2005; Xue et al. 2008; Chasse & Miller 2010; Incze et al. 2010). Larval behaviour may also have a significant influence on the population structuring of the shore crab, *Carcinus maenas*, with genetic differentiation detected between Sweden and the UK despite the absence of barriers to gene flow (Domingues et al. 2010).

Regardless of the behavioural uncertainties in *H. gammarus*, it is evident from the PTMs and the genetic data that there is high connectivity between Lundy NTZ and

the Devon coast, particularly if larval development is completed within three weeks (stage III). Water temperature is considered a controlling factor in *H. americanus* larval development, with development time decreasing with increasing water temperature (MacKenzie 1988). Therefore, a three week development is feasible for *H. gammarus* under warmer conditions. This high connectivity with Devon is likely to have a positive impact on local fisheries. Conversely, the prospect of high settlement along the coasts of Ireland by surface larvae (PTMs 2-3) highlights the possibility of long distance population connectivity, for which there is support from the genetic data.

Self-recruitment by Lundy NTZ lobsters was low under all behavioural strategies, with the highest settlement potential being 5.33% at the end of stage I during PTM-4, though the likelihood of achieving this is minimal, since water temperatures would be unable to support larval development in less than seven days. However, an increase in dispersal time sees a decrease in self-recruitment at Lundy NTZ, with only 1.37% settlement potential by the end of stage IV. Therefore, the low levels of self-recruitment infer that the majority of larvae supporting Lundy population persistence are sourced from elsewhere, which is compatible with the overall model of an effectively panmictic population within the studied area. The particle back-tracking models suggest that larvae mainly originate from the east, especially the North Devon coast and the Severn Estuary, with Lundy NTZ representing a 'sink' for larval recruits.

Significant deviations from Hardy-Weinberg equilibrium, specifically heterozygote deficits (non-zero values of F_{IS}) are a common feature among marine invertebrates (Addison & Hart 2005). However, an interesting feature of the genetic data here was the significantly higher values of F_{IS} (and corrected relatedness) obtained for the NTZ sample compared to all other samples. Such a pattern could be generated by factors such as inbreeding and selection. Furthermore, the large census population sizes of *H. gammarus*, particularly at the NTZ (Wootton et al. 2012) and comparable levels of genetic polymorphism makes inbreeding an unlikely factor. Finally, as the pattern was replicated across loci (with all loci assumed to be neutral)

selection against heterozygotes can also be considered as extremely unlikely. Therefore, other explanations must be sought.

Significant heterozygote deficits are often attributed to Wahlund effects (i.e. the inadvertent sampling of multiple subpopulations within a single sample). Based on the low level of genetic structure throughout the studied region, a spatial Wahlund effect within the NTZ can be discounted. However, for many highly fecund marine species with pelagic larval stages, large variances in reproductive success (Hedgecock 1994) can serve to generate genetic differences between cohorts (cohort Wahlund) (Ruzzante et al. 1996) within a panmictic population. The elevated F_{IS} and relatedness estimates could, therefore, reflect a more pronounced cohort Wahlund effect detected at the NTZ due to the presence and sampling of a greater number of distinct cohort classes compared to other sites (where older cohorts may be removed by harvesting). However, the observed patterns for the NTZ may also reflect changes to 'within cohort' recruitment patterns. Wootton et al. (2012) reported that increased lobster densities at the NTZ were associated with negative effects such as increased shell disease and injuries. Likewise, Davies et al. (2014) stated that intraspecific competition, due to more abundant and larger lobsters in the NTZ, may be responsible for the increased prevalence of disease in the local population. Such effects could serve to increase the variance in reproductive success among individuals through either stochastic and/or selective processes (Planes & Lenfant 2002). Furthermore, while biophysical modelling indicated very low levels of self-recruitment at the NTZ, the observed patterns could also be compatible with a reduced proportion of migrants surviving to adult stages at the NTZ compared to other sites.

The results reported here indicate that while NTZs may serve to increase census population sizes, they may also contribute to a reduction in N_e. Many highly fecund marine taxa are characterised by low N_e/N_c ratios (Hauser et al. 2002; Hoarau et al. 2005), which has important implications for population persistence on ecological and evolutionary timescales. Future genetic analysis of distinct cohorts is needed to provide insight into the relative roles of 'within cohort' and 'between cohort' recruitment dynamics shaping the observed patterns for the NTZ.

Whilst the model predicts high connectivity between Lundy NTZ and Devon, self-recruitment is actually very low, suggesting that the larvae supporting Lundy must be sourced from elsewhere. Therefore, management should be on the scale of the area as a whole to ensure the sustainable management of these 'sources'. Accordingly, the monitoring or protection of suitable spawning habitat and nursery grounds would provide the early life stages of *H. gammarus* an increased chance of survival. A ban on landing ovigerous 'berried' females, in addition to the enforcement of the existing MLS, would enable the lobsters to reproduce, at least once, before being landed. Furthermore, it is crucial that the highlighted sources of larvae (e.g. North Devon and the Severn Estuary) are offered some level of protection. These proposed measures would contribute towards the effective management of *H. gammarus* within the Irish Sea.

3. MICROSATELLITE ANALYSIS OF POPULATION STRUCTURE OF THE EDIBLE CRAB, *CANCER PAGURUS*, AT DIFFERENT GEOGRAPHICAL SCALES

3.1. INTRODUCTION

Fishing industries in the UK are shifting focus from fin-fisheries and turning towards exploiting shellfisheries resources (Molfese et al. 2014). However, the lack of baseline data on shellfish stock structuring has led to concerns about the conservation and management of the target species (Anderson et al. 2011; Howarth et al. 2013), emphasising the need for structure and connectivity of populations to be empirically assessed. Genetic markers can describe population structure by defining patterns of variation at neutral genetic loci that have been determined by the opposing evolutionary forces of gene flow, genetic drift and mutation (Utter 1991; Hellberg et al. 2002). Knowing this information should contribute towards effective fisheries management (Ward 2000).

Consequently, there has been considerable research interest in the study of population genetics of marine decapods, particularly those of high fishery value. Since decapods are characterised by large population sizes and wide geographical distributions, as well as high dispersal potential due to adult movements and/or long pelagic larval phases (Bennett & Brown 1983; Eaton et al. 2003), a low level of population structure is assumed. Such a pattern has been reported for a number of species including the spiny spider crab, *Maja brachydactyla*, with the detection of weak geographical structure in the Northeast Atlantic (Sotelo et al. 2008). However, complex patterns of genetic differentiation have been discovered at various geographic scales in several commercial species, indicating failure in dispersal potential: spider crabs *Inachus dorsettensis* and *Hyas coarctatus* (Weber et al. 2000), European lobster *Homarus gammarus* (Jørstad et al. 2004; Jørstad et al. 2005; Triantafyllidis et al. 2005), Norway lobster *Nephrops norvegicus* (Stamatis et al. 2004; Stamatis et al. 2006), and the coastal shrimp *Crangon crangon* (Weetman et al. 2007). For instance, significant structuring was identified across the European

distribution range of *H. gammarus*, with distinct genetic clusters in the Mediterranean, northern Norway, Netherlands and remaining Atlantic samples using mitochondrial DNA (mtDNA) (Triantafyllidis et al. 2005) and allozymes (Jørstad et al. 2005). Similarly, population differentiation was revealed in Northeast Atlantic and Mediterranean populations of *N. norvegicus*, using mtDNA (Stamatis et al. 2004) and allozymes (Stamatis et al. 2006). Genetic structuring was observed in the commercially important *C. crangon*, with clear groupings in western Britain, the eastern English Channel and the Baltic Sea, as identified by Amplified Fragment Length Polymorphism (AFLP) analysis (Weetman et al. 2007). A mtDNA study found significant differences between samples of *H. gammarus* that were collected from sites separated by a coastal distance of 142 km in northern Norway (Jørstad et al. 2004). On an even finer scale, allozymes revealed significant genetic differentiation between samples of *I. dorsettensis* and *H. coarctatus*, over a geographical distance of only 40 km (Weber et al. 2000). These deviations from predictions of widescale panmixia may be driven by features that are intrinsic and/or extrinsic to certain taxa in each case. For instance, it was likely that local hydrological conditions preventing larval dispersal caused geographical isolation of the *H. gammarus* populations (Jørstad et al. 2004), with gene flow being restricted by distance and hydrographic features in *C. crangon* (Weetman et al. 2007). *N. norvegicus* are said to have experienced a recent population expansion (Stamatis et al. 2004), whereas the low degree of differentiation in the remaining Atlantic samples of *H. gammarus* were thought to reflect the recent postglacial establishment of populations from a common refuge (Triantafyllidis et al. 2005). Evidently, the complex interaction of numerous factors determines the genetic structure and connectivity in marine species. Yet they cannot account for all patterns observed (Weber et al. 2000), indicating the need for greater research effort in this multifaceted field.

The edible crab, *Cancer pagurus*, is widely distributed in the Northeast Atlantic, with its range extending from northern Norway to northwest Africa (Christiansen 1969). Its landings contribute nearly 20% of the total shellfish landings in the UK, with 32,111 tonnes landed in 2012 (Cefas 2013). That year, a total of 162,754 tonnes of shellfish were landed worth £300.801 million, of which crab represented almost

14% with £41.67 million generated (Cefas 2013), highlighting the economic importance of the species to the UK fishing industry.

It is known from tagging studies that adult female *C. pagurus* have the ability to travel long distances, with one female being recaptured 302.4 km from the point of release (Hunter et al. 2013). The phenomenon of female migrations occurs throughout the Northeast Atlantic with reports in the English Channel (Bennett & Brown 1983; Hunter et al. 2013), the Celtic Sea (Fahy & Carroll 2008), the North Sea (Edwards 1979) and the Skagerrak and Kattegat (Ungfors et al. 2007). Tagging studies have mainly focused on the English Channel, where a predominantly westward migration has been demonstrated, with female *C. pagurus* in the eastern Channel migrating further than those in the western Channel (Hunter et al. 2013). This movement had previously been observed through a mark-recapture experiment focusing on the English Channel in 1968-1974 (Bennett & Brown 1983). The recurring westward female migrations in this area, suggest that the behaviour is somewhat beneficial to the species and, consequently, has been linked to brooding (Hunter et al. 2013). Females with developing gonads move into deeper waters, where the tidal flows are low, to ensure a suitable spawning location; soft sand or gravel enable the formation of a hollow in which the abdomen can be lowered to ensure the attachment of the eggs to the pleopods (Edwards 1979). It has been estimated that brooding grounds occur at various locations throughout the English Channel, with mature *C. pagurus* not showing site fidelity between years (Hunter et al. 2013). Previously, the westward migrations have been considered as an example of counter-current spawning behaviour (Pawson 1995; Eaton et al. 2003). Females can migrate significantly longer distances, 1.8-8.4 times farther, more frequently than males (Ungfors et al. 2007), with male movement considered to be the result of nomadism (Bennett & Brown 1983).

C. pagurus is highly fecund with 0.5 to 2.9 million eggs produced per female (Edwards 1979; Ungfors 2007). Larval development occurs within the temperature range of 14 °C ± 3 °C (Weiss et al. 2009), with the planktotrophic larvae remaining pelagic for approximately three months (Eaton et al. 2003; Hunter et al. 2013). Consequently, the highly mobile adults, particularly the females, and the long larval

phase both contribute towards the large dispersal potential of *C. pagurus* and has led to the assumption of high gene flow around the British coast.

Due to the economic importance of the species, genetic research is being conducted to gain an understanding of the size and demography of *C. pagurus* populations. Thus far, studies have focused on British waters (Shaw 2003), Irish waters (Moran 2009) and the Kattegat-Skagerrak area (Ungfors et al. 2009), using three, six and eight microsatellite markers respectively. Significant structuring, not following a particular geographical pattern, was detected in the UK crab population using three microsatellites, with low but significant genetic differences between almost all sites (Shaw 2003). Similarly, significant differentiation was found between samples approximately 55 km apart at Galway Bay, Ireland, whilst distant sites appeared connected (Moran 2009). Finally, large-scale genetic mixing was observed in the Kattegat-Skagerrak area, with a lack of spatial and temporal genetic differentiation (Ungfors et al. 2009). Evidently, there appears to be varying degrees of differentiation in *C. pagurus* throughout its range, highlighting the need for management at different geographical scales throughout its distribution. There is a demand for studies incorporating both fine and regional scale sampling to understand the biological significance of genetic patterns.

The aim of this study was to investigate the spatial, genetic structure on a fine-scale of *C. pagurus* in the Irish Sea, using microsatellite markers. The data generated were then collated with genotypic data collected across the NE Atlantic in order to examine the population structure of *C. pagurus* on a wider-scale. It is hoped that the knowledge gained from this work can be used to better inform future stock management.

3.2. MATERIALS AND METHODS

3.2.1 Study Area

Haemolymph was collected from a total of 514 crabs at 13 locations within the Irish Sea (Table 3.1, Figure 3.1). Sampling locations were along the coasts of Wales [West Anglesey (AW), Menai Straits (NW), Aberystwyth (CB), Oxwich (OX), Mumbles (M) and the Gower (SW)] and the Republic of Ireland [Howth (ND), Dun Laoghaire (SD), Carne (WEX) and Dunmore East (WF)]. Samples were also collected at Lundy Island in the Bristol Channel [Lundy NTZ (LNTZ), outside the NTZ in May (LIA) and in June (LIB)]. Haemolymph was collected from commercially caught individuals, hence above the 130 mm MLS for the Irish Sea region (SEAFISH 2013), except for NW, OX and M where juveniles were sampled.

Table 3.1. The location, assigned code and coordinates of the sampling locations in the Irish Sea dataset, as well as the date of sample collection (S) and the number of individuals sampled (N). The sex of the individual was recorded when possible (M : F).

Location	Code	Coordinates	S	N	M : F
Howth, North Dublin (IE)	ND	53.469° N, 6.084° W	Jul-11	48	16 : 32
Dun Laoghaire, South Dublin (IE)	SD	53.217° N, 6.085° W	Jul-11	48	17 : 31
Carne, Wexford (IE)	WEX	52.184° N, 6.302° W	Jul-11	46	9 : 37
Dunmore East, Waterford (IE)	WF	52.085° N, 7.033° W	Jul-11	47	7 : 40
West Anglesey, North Wales (UK)	AW	53.213° N, 4.606° W	Apr-10	15	13 : 1
Menai Straits, Anglesey (UK)	NW	53.220° N, 4.163° W	Jan-11	48	30 : 18
Aberystwyth, Cardigan Bay (UK)	CB	52.415° N, 4.236° W	Oct-10	48	27 : 21
Oxwich, South Wales (UK)	OX	51.566° N, 4.147° W	Jan-Jun-11	48	28 : 20
Mumbles, South Wales (UK)	M	51.570° N, 3.980° W	Feb-Jun-11	48	28 : 20
Gower, South Wales (UK)	SW	51.550° N, 4.144° W	Jun-10	16	14 : 2
Lundy Island (outside NTZ) (UK)	LIA	51.205° N, 4.682° W	May-10	47	35 : 11
Lundy Island (outside NTZ) (UK)	LIB	51.205° N, 4.682° W	Jun-10	22	17 : 5
Lundy Island NTZ (UK)	LNTZ	51.189° N, 4.649° W	May-10	33	27 : 6

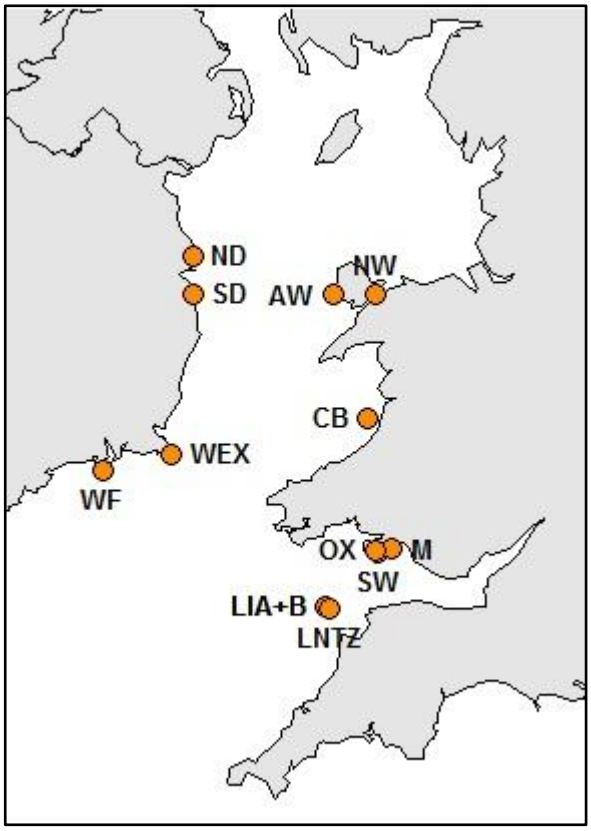


Figure 3.1. Sampling locations of the edible crab *Cancer pagurus* in the Irish Sea.

3.2.2 DNA Extraction and Microsatellite Genotyping

Haemolymph was taken using a 2 ml Terumo syringe with a G23x25 mm needle (VWR International Ltd.) and preserved in absolute ethanol (1:8). DNA was extracted using either the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, CA-USA) or the CHELEX 100 (Bio-Rad, CA-USA) protocol (Walsh et al. 1991; Goff & Moon 1993). The Animal Blood Spin-Column Protocol from the QIAGEN kit was modified to include an initial centrifugation step of 400 µl of the haemolymph/ethanol mixture (sample) for 5 minutes at 7000 x g in order to precipitate the haemolymph cells into a pellet easily separated from the alcoholic fraction.

Twelve species-specific microsatellite loci developed by (McKeown & Shaw 2008a) were amplified in a single multiplex polymerase chain reaction (PCR) (Figure 3.2). Amplification was carried out using a QIAGEN Multiplex PCR Kit (QIAGEN, CA-USA) in a final volume of 15 µl, containing 7.5 µl of Multiplex Kit Buffer and 1 µl of

genomic DNA. Primer volumes varied: 0.35 µl Cpag-4C1; 0.3 µl Cpag-1B9, Cpag-3A2, Cpag-3D7, Cpag-4 and Cpag-5D8; 0.25 µl Cpag-38; 0.2 µl Cpag-1C8, Cpag-2D7, Cpag-6C4B and Cpag-15; and 0.1 µl Cpag-2A5B. The PCR cycle involved an initial denaturation step at 95 °C for 15 minutes, followed by 34 cycles of 45 seconds at 94 °C, 45 seconds at 55 °C and 45 seconds at 72 °C, and a final extension step at 72 °C for 45 minutes. Products were then run on an ABI 3730 Genetic Analyzer (Applied Biosystems) alongside a GS500LIZ size standard and alleles were scored using GeneMapper 4.0 (Applied Biosystems).

The amplification of locus Cpag-2D7 failed to provide good quality data, hence this locus was removed from the analysis.

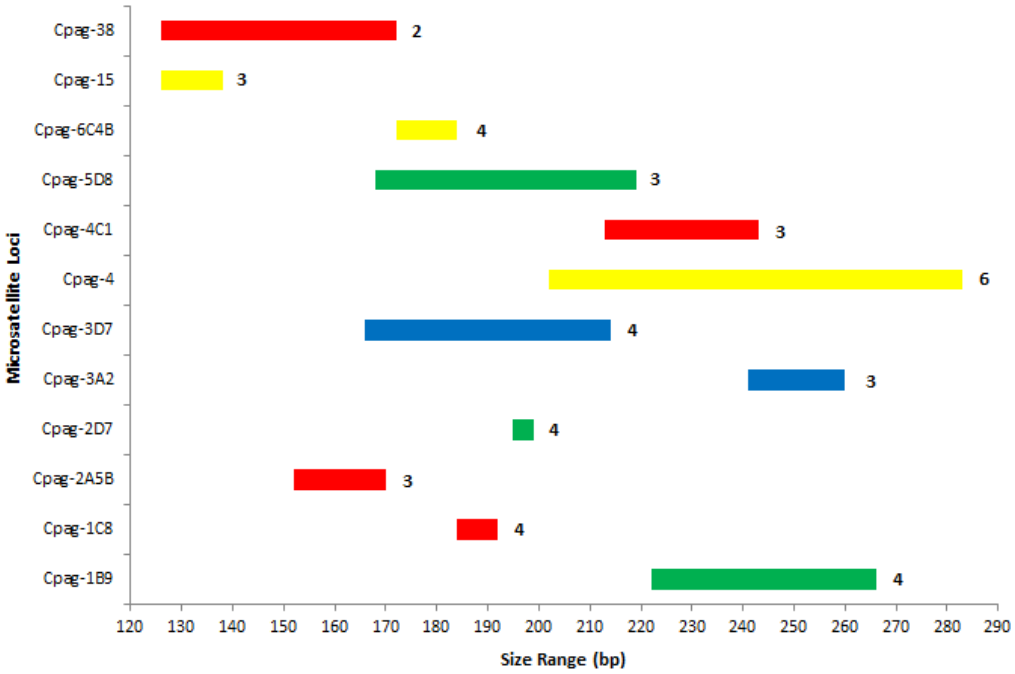


Figure 3.2. Multiplex plan including all twelve microsatellite loci. The colours represent the fluorescent dye chosen for the forward primer: blue – 6FAM, green – VIC, yellow – NED, and red – PET (Applied Biosystems). The bold number adjacent to the bar indicates the size of the repetitive unit.

3.2.3. Expanded Study of Northeast Atlantic Populations of *Cancer pagurus*

The dataset generated in the present study was then merged with the data available from a previous study in order to expand the geographical range taken into consideration so as to investigate patterns of structure at a wider spatial scale.

This additional dataset comprised of 919 *C. pagurus* from across the NE Atlantic screened for 8 loci (Cpag-5D8, Cpag-4, Cpag-6C4B, Cpag-3A2, Cpag-1B9, Cpag-3D7, Cpag-2A5B and Cpag-15). Sampling locations were scattered across the species distribution range [from Ireland: Donegal (NNW), Waterford (NSE) and Galway Bay (NSW); Wales, Aberystwyth (NAB); the Celtic Sea (Pendeen); the English Channel: Brittany (7Brt), Jersey (6Jer) and Hastings (6Has); the North Sea: Harwich (6Har) and Northumberland (6Sea); the Norwegian Sea (NNO)] (Table 3.2, Figure 3.3).

To calibrate the datasets, 96 randomly chosen samples from the added NE Atlantic samples were re-amplified and genotyped, and then compared with the genotypes of samples specifically collected for this study. The final dataset included a total of 1433 individuals from 24 locations across the NE Atlantic screened at 8 loci.

Table 3.2. The location, assigned code and coordinates of the additional sampling locations in the NE Atlantic dataset, as well as the date of sample collection (S) and the number of individuals sampled (N).

Location	Code	Coordinates	S	N
Donegal, NW Ireland (IE)	NNW	54.558° N, 8.315° W	Jul-07	30
Waterford, SE Ireland (IE)	NSE	52.085° N, 7.033° W	Jul-07	31
Galway Bay, Ireland (IE)	NSW	53.196° N, 9.279° W	Jul-07	46
Aberystwyth, Cardigan Bay (UK)	NAB	52.415° N, 4.236° W	Aug-00	69
Pendeen, Celtic Sea (UK)	Pen	50.153° N, 5.740° W	Jun-06	102
Brittany, English Channel (FR)	7Brt	47.250° N, 5.500° W	Jul-06	102
Jersey, English Channel (UK)	6Jer	49.117° N, 2.233° W	Sep-07	84
Hastings, English Channel (UK)	6Has	50.718° N, 0.662° E	Oct-06	162
Harwich, North Sea (UK)	6Har	51.590° N, 1.590° E	May-05	159
Northumberland, North Sea (UK)	6Sea	55.330° N, 1.320° E	Sep-05	65
Norwegian Sea (NO)	NNO	62.400° N, 6.390° E	Dec-04	69

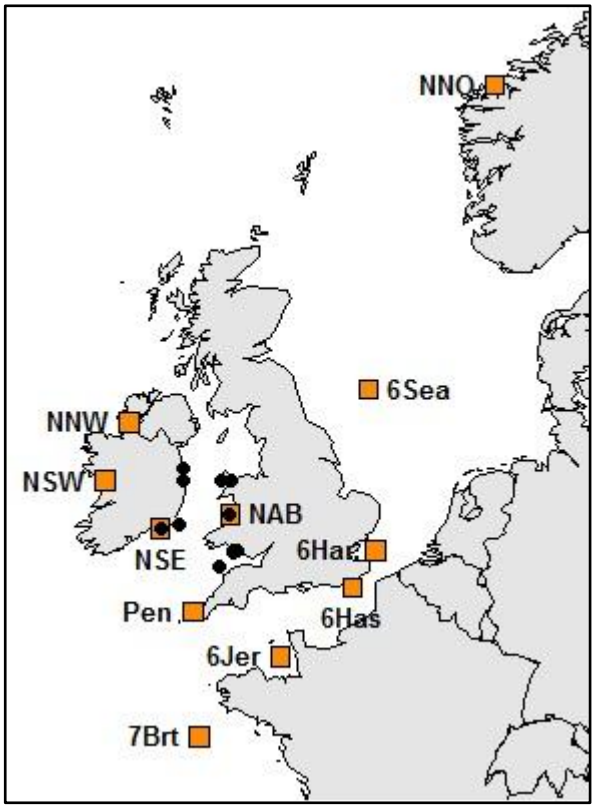


Figure 3.3. Sampling locations of the edible crab *Cancer pagurus* added from N.J. McKeown (orange squares). The black circles are the Irish Sea dataset locations collected in the present study.

3.2.4. Statistical Analysis

General descriptive statistics and measures of variability (H_e, H_o, N_A, A_R, F_{IS}), plus tests for Hardy-Weinberg conformance, null alleles and selective neutrality were performed using the methods outlined in Section 2.2.4.1. In order to assess the statistical power when testing for genetic differentiation, power analysis was carried out using the program POWSIM (Ryman & Palm 2006) (Section 2.2.4.2). Population structuring was then investigated using the Bayesian clustering method implemented in STRUCTURE 2.3 (Pritchard et al. 2000; Falush et al. 2003, 2007), making no assumptions about location (Section 2.2.4.3). Global and pairwise F_{ST} were assessed in FSTAT 2.9.3 (Goudet 1995) and the more recently proposed genetic diversity indices G'_{ST} (Hedrick 2005) and D_{est} (Jost 2008) were estimated using the web-based application SMOGD (Crawford 2010) (Section 2.2.4.4). PCoA was performed in R 3.0.2 (R Core Team 2013) to produce comparative MDS plots of

D_{est} and F_{ST}, with the correlation between the two genetic distance matrices measured with a Mantel test (Mantel 1967; Smouse et al. 1986), using the R package ECODIST (Goslee & Urban 2007). An analysis of molecular variance (AMOVA) was performed in ARLEQUIN to evaluate the amount of population genetic structure between samples from the east of the Irish Sea (the 'Wales' grouping) and those to the west of the Irish Sea (the 'Ireland' grouping). IBD was tested using IBDWS (Jensen et al. 2005) (Section 2.2.4.4). N_e was estimated using LDNE (Waples & Do 2008), with evidence of population size reductions assessed using BOTTLNECK 1.2 (Piry et al. 1999) and the M-ratio between the total number of alleles and the overall range in allele size (Garza & Williamson 2001); see Section 2.2.4.6 for the theory behind these methods. Finally, the variation among groups of samples for measures of A_R, H_O, H_S, F_{IS}, F_{ST} and relatedness was tested using the permutation approach in FSTAT (Section 2.2.4.7). Two different groupings were analysed: (1) males versus females across all populations to test for potential sex differences, and (2) Lundy NTZ versus the other sampled sites in the Irish Sea to test for site-associated differences. Sex differences could only be assessed for the Irish Sea dataset for which sex was recorded.

3.3. RESULTS

3.3.1. Genetic Diversity in the Irish Sea

H_e ranged from 0.577 in the West Anglesey (AW) population to 0.641 in the Lundy Island (LIA) population. H_o ranged from 0.533 for Lundy Island NTZ (LNTZ) to 0.608 for the Gower (SW) population. LIA had the highest N_A across loci (9.36), whereas AW had the lowest (5.73). Furthermore, AW had the lowest A_R value with 5.73. In contrast, the population with the highest allelic richness was SW with 6.80. Two populations were shown to have a significant departure from HWE: LIA (F_{IS} = 0.140; 95% CI: 0.054 – 0.216) and LNTZ (F_{IS} = 0.132; 95% CI: 0.032 – 0.196) (Table 3.3). Moreover, locus Cpag-4 exhibited deviations from HWE at 6 out of the 13 populations tested ($P < 0.05$). For the remaining loci, no locus showed deviations from HWE in more than 3 populations (Table 3.4). In order to establish whether Cpag-4 was having an effect, the locus was removed and F_{IS} was recalculated; there were no changes in the levels of significance, with both LIA (F_{IS} = 0.145, 95% CI: 0.04848 - 0.22691) and LNTZ (F_{IS} = 0.133, 95% CI: 0.02664 - 0.19931) remaining significant. There was no evidence for linkage disequilibrium between any locus pair across all populations (Appendix 5).

FreeNA detected a moderate likelihood ($0.05 \leq r < 0.20$) of null alleles, although not consistently across loci or populations, and none of the loci were shown to have a large chance ($r \geq 0.20$) of harbouring them (Table 3.5). Cpag-4 was only found to have a moderate likelihood of null alleles at two out of the 13 populations: NW (r = 0.058) and WF (r = 0.057).

LOSITAN confirmed the neutrality of the 11 microsatellite markers that were used, with none being selected as candidates for balancing or positive selection (Figure 3.4).

POWSIM indicated that the data (loci and average sample size) conferred a high probability of even subtle differentiation between samples (Table 3.6).

Table 3.3. Genetic diversity parameters inferred from microsatellite markers for *Cancer pagurus* in the Irish Sea. H_e and H_o expected and observed heterozygosities, N_A number of alleles, A_R allelic richness, and F_{IS} inbreeding coefficient. Significant values are marked with an asterisk (95% CI).

	H_e	H_o	N_A	A_R	F_{IS}
AW	0.577	0.552	5.73	5.73	0.046
CB	0.614	0.585	8.73	6.34	0.047
LIA	0.641	0.552	9.36	6.67	0.140*
LIB	0.602	0.583	7.64	6.54	0.033
LNTZ	0.613	0.533	7.64	5.99	0.132*
NW	0.614	0.578	8.73	6.19	0.059
SW	0.615	0.608	7.00	6.80	0.011
ND	0.612	0.591	9.09	6.42	0.035
SD	0.591	0.572	9.09	6.33	0.032
WF	0.601	0.578	8.73	5.97	0.039
WEX	0.628	0.591	8.18	6.24	0.060
M	0.593	0.576	8.09	5.97	0.028
OX	0.614	0.604	9.27	6.51	0.017

Table 3.4. Hardy-Weinberg probability test for *Cancer pagurus* in the Irish Sea. Significant values are marked with an asterisk ($P < 0.05$).

PHWE	Cpag 4	Cpag 1B9	Cpag 1C8	Cpag 2A5B	Cpag 3A2	Cpag 3D7	Cpag 4C1	Cpag 5D8	Cpag 6C4B	Cpag 15	Cpag 38
AW	0.108	0.703	-	1.000	0.731	0.726	1.000	0.344	0.340	1.000	0.487
CB	**0.008***	0.321	-	0.777	**0.013***	0.280	0.199	0.148	0.166	0.664	0.241
LIA	**0.000***	0.598	**0.002***	0.298	0.828	**0.005***	**0.020***	**0.000***	0.732	0.252	**0.017***
LIB	1.000	0.566	-	0.052	0.700	0.666	0.207	0.132	0.942	0.710	0.535
LNTZ	**0.000***	0.209	-	0.669	0.080	0.978	0.150	0.288	0.085	0.264	0.582
NW	**0.016***	0.662	1.000	0.982	0.723	0.873	0.146	0.257	0.182	0.654	0.419
SW	1.000	**0.018***	**0.032***	0.616	0.871	0.601	0.513	0.249	0.312	0.193	0.956
ND	0.836	0.826	1.000	0.881	0.866	0.133	0.198	0.830	0.793	0.848	0.606
SD	0.132	0.603	1.000	0.582	0.409	0.559	0.400	0.456	0.911	1.000	0.408
WF	**0.002***	0.455	1.000	0.294	0.386	0.725	0.438	0.608	0.738	0.727	0.669
WEX	**0.000***	0.449	**0.033***	0.757	0.817	0.423	**0.016***	0.996	0.991	0.738	**0.022***
M	0.747	0.065	-	0.844	0.876	0.538	**0.014***	0.489	0.085	0.486	0.489
OX	0.357	0.878	1.000	0.101	0.836	0.682	0.562	**0.017***	0.287	**0.025***	0.150

Table 3.5. The estimate of null allele frequency per locus per population of *Cancer pagurus* in the Irish Sea using FreeNA (Bootstrap = 1000). The values in white squares have a negligible chance of containing null alleles (r < 0.05) and those in grey squares have a moderate chance (0.05 ≤ r < 0.20).

	Sampling Locations												
Loci	**AW**	**CB**	**LIA**	**LIB**	**LNTZ**	**NW**	**SW**	**ND**	**SD**	**WF**	**WEX**	**M**	**OX**
Cpag 4	0.022	0.000	0.046	0.000	0.049	0.058	0.000	0.000	0.017	0.057	0.047	0.000	0.007
Cpag 1B9	0.000	0.007	0.000	0.000	0.063	0.036	0.095	0.000	0.000	0.000	0.017	0.049	0.009
Cpag 1C8	0.001	0.000	0.123	0.000	0.000	0.000	0.165	0.000	0.000	0.000	0.092	0.001	0.000
Cpag 2A5B	0.017	0.003	0.057	0.128	0.042	0.000	0.000	0.000	0.025	0.000	0.000	0.017	0.000
Cpag 3A2	0.000	0.069	0.031	0.043	0.119	0.000	0.017	0.000	0.000	0.000	0.000	0.000	0.000
Cpag 3D7	0.011	0.056	0.111	0.000	0.034	0.000	0.000	0.066	0.039	0.033	0.080	0.016	0.000
Cpag 4C1	0.010	0.067	0.079	0.000	0.099	0.000	0.000	0.000	0.013	0.000	0.061	0.073	0.000
Cpag 5D8	0.024	0.017	0.091	0.045	0.034	0.041	0.059	0.007	0.033	0.004	0.000	0.007	0.091
Cpag 6C4B	0.000	0.013	0.032	0.000	0.000	0.026	0.000	0.000	0.000	0.045	0.000	0.000	0.000
Cpag 15	0.000	0.000	0.017	0.000	0.003	0.047	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Cpag 38	0.000	0.000	0.067	0.000	0.000	0.033	0.000	0.049	0.034	0.007	0.064	0.000	0.000

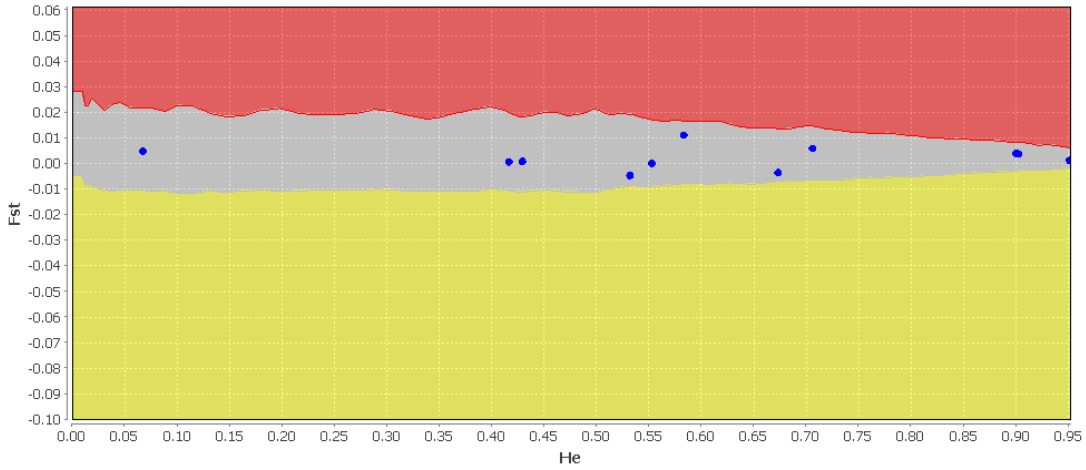


Figure 3.4. LOSITAN neutrality test for the 11 microsatellite makers (blue dots) used on *Cancer pagurus* in the Irish Sea. Markers are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area).

Table 3.6. Results of POWSIM analysis indicating the probability of detecting differentiation by exact tests at various levels of simulated true structuring (F_{ST} 0.001-0.05). Proportion of significant tests at F_{ST} = 0 indicate the Type I error probability.

F_{ST}							
0	**0.001**	**0.0025**	**0.005**	**0.01**	**0.02**	**0.025**	**0.05**
0.05	0.08	0.29	0.66	0.98	1	1	1

Overall, there was no evidence of any population structuring, since K=1 with a mean likelihood of LnP(K) = -16219.88 (Figure 3.5). Furthermore, there was an equal assignment of each individual to K clusters (Figure 3.6).

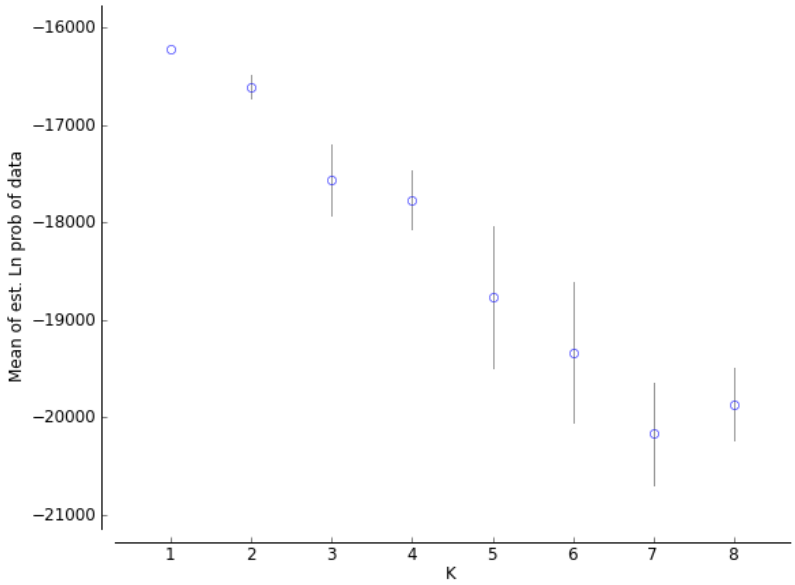


Figure 3.5. Plot of mean likelihood L(K) and variance per K value from STRUCTURE on the Irish Sea dataset containing 514 individuals genotyped for 11 microsatellite loci.

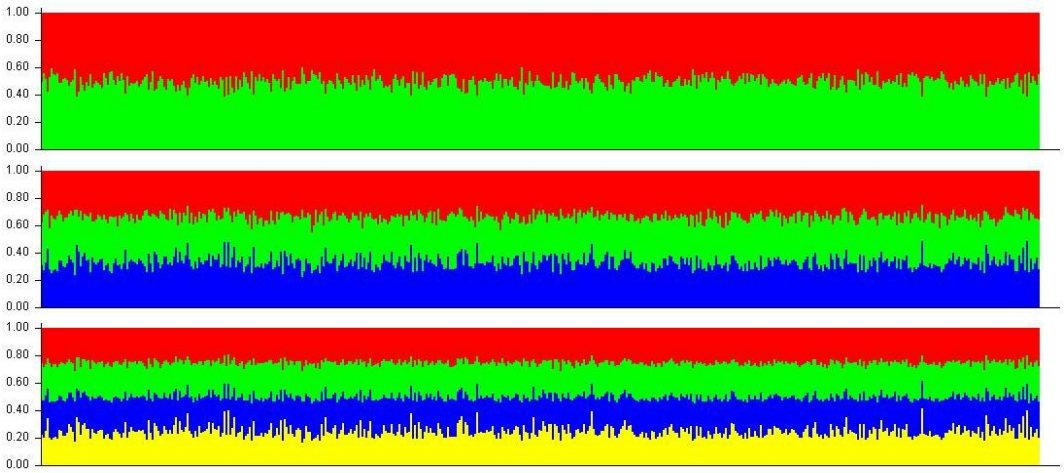


Figure 3.6. STRUCTURE bar plots for *Cancer pagurus* in the Irish Sea for K = 2, K = 3 and K = 4 for 514 individuals showing assignment to K clusters. Each vertical bar represents an individual and each colour a different cluster.

Global F_{ST} was non-significant at 0.001 (95% CI: -0.000 – 0.003). When looking at the pairwise F_{ST} values, all of the Irish Sea populations appear to be similar to each other, with no significant differences being detected after Bonferroni correction (Table 3.7). When ENA correction was applied to the global and pairwise F_{ST} values there were changes in the level of significance. For instance, global F_{ST} before ENA correction was 0.001099 (95% CI: -0.000431 - 0.002719) and was 0.002272 (95% CI: 0.000490 - 0.004442) after ENA correction, thus becoming significant although remaining very low and of the same order of magnitude. In addition, three pairwise population comparisons became significant after ENA correction: LIA and SD; LIA and OX; and LNTZ and SW (Appendix 6). The MDS plot of the ENA-corrected F_{ST} values revealed a similar pattern to the non-corrected F_{ST} values (ENA corrected F_{ST} – Figure 3.7, F_{ST} – Figure 3.8).

Table 3.7. F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Cancer pagurus* in the Irish Sea. Here, the indicative adjusted nominal level (5%) for multiple comparisons has been calculated as 0.000641 (Bonferroni correction). Significant values at the $P < 0.05$ level are marked in bold.

	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX
AW		0.156	0.891	0.578	0.331	0.145	0.188	0.437	0.422	0.216	0.171	0.215	0.330
CB	0.010		0.250	0.915	0.079	0.516	**0.049**	0.401	0.629	0.521	0.146	0.363	0.153
LIA	-0.003	0.003		0.929	0.104	0.728	0.266	0.482	0.263	0.512	0.270	**0.034**	0.112
LIB	-0.001	-0.006	-0.003		0.938	0.950	0.569	0.928	0.978	0.963	0.866	0.865	0.900
LNTZ	0.012	0.006	0.002	-0.004		0.708	0.055	0.060	0.143	0.139	0.163	**0.036**	0.251
NW	0.015	-0.001	0.002	-0.005	0.000		0.210	0.726	**0.035**	0.844	0.300	0.274	0.824
SW	0.009	**0.001**	-0.001	0.000	0.007	0.004		0.561	0.215	0.060	**0.028**	**0.015**	0.344
ND	0.011	0.003	-0.001	-0.002	0.004	0.002	-0.003		0.949	0.608	0.187	0.104	0.788
SD	0.004	0.004	0.005	-0.001	0.007	**0.010**	0.002	0.000		0.584	**0.042**	0.367	0.674
WF	0.008	-0.002	0.003	-0.006	0.006	0.000	0.004	-0.001	0.001		0.592	0.431	0.666
WEX	0.006	-0.001	-0.002	-0.006	-0.002	-0.001	**-0.003**	0.001	**0.005**	0.000		**0.031**	0.065
M	0.006	-0.003	**0.002**	-0.006	**0.003**	0.002	**0.002**	-0.002	-0.001	-0.003	**0.001**		0.096
OX	0.006	0.001	0.003	-0.004	0.005	0.002	-0.004	-0.001	-0.002	-0.002	0.001	0.000	

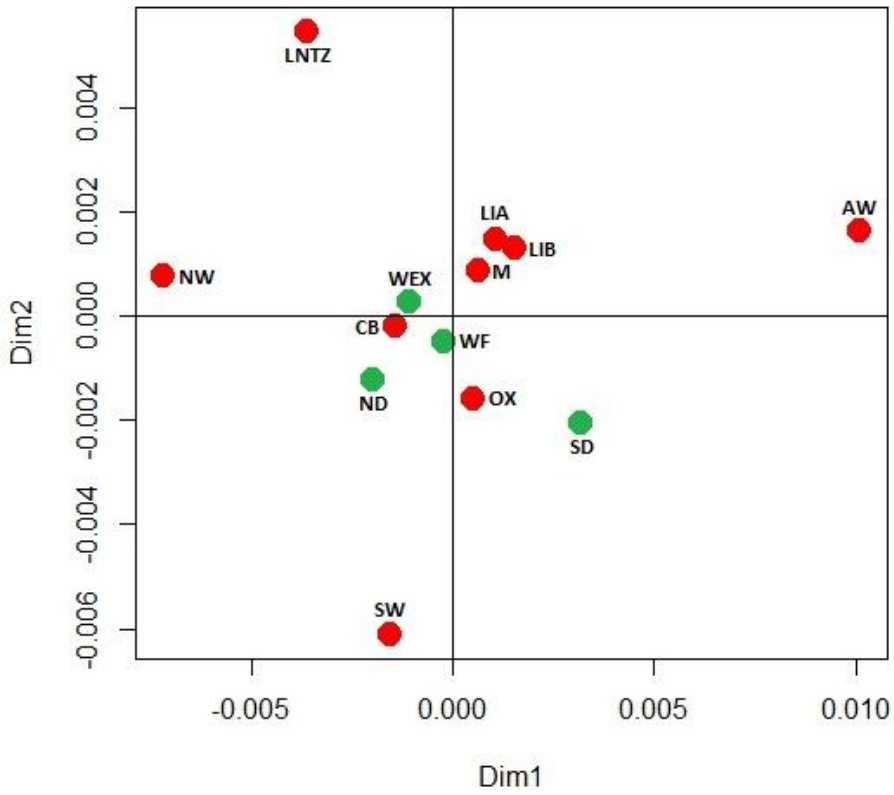


Figure 3.7. MDS plot of ENA-corrected pairwise F_{ST} values of *Cancer pagurus* in the Irish Sea. Red circles are eastern Irish Sea localities and green denotes western Irish Sea localities.

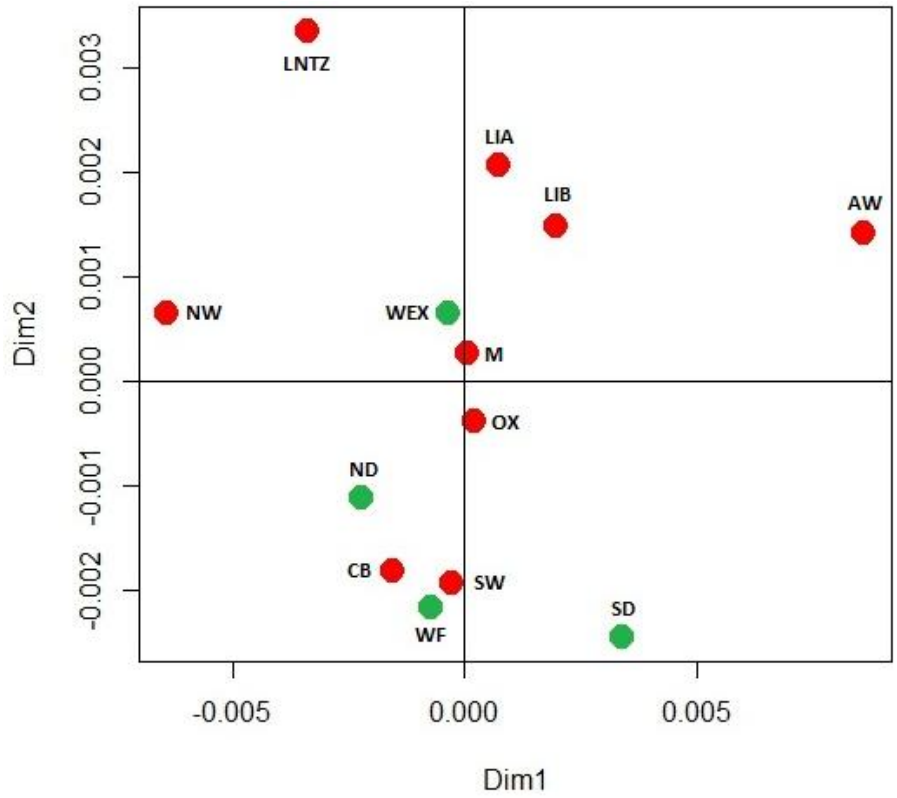


Figure 3.8. MDS plot of pairwise F_{st} values of *Cancer pagurus* in the Irish Sea. Red circles are eastern Irish Sea localities and green denotes western Irish Sea localities.

Locus-specific G'_{ST} and D_{est} values were very low, with Cpag-5D8 shown to have the most genetic differentiation in both measures but with only < 4% of the maximum possible (G'_{ST} = 0.040, D_{est} = 0.036). Cpag-1B9, Cpag-2A5B and Cpag-6C4B revealed no genetic differentiation in either measure. Furthermore, Cpag-15 and Cpag-1C8 had a D_{est} value of zero (Table 3.8). Pairwise D_{est} revealed low levels of differentiation that were similar to those shown by F_{ST} (r = 0.696, P < 0.001) (Table 3.9) and the MDS plots from each measure revealed no geographical pattern of sample clustering (F_{ST} – Figure 3.8, D_{est} – Figure 3.9).

Table 3.8. Locus-specific G'_{ST} and D_{est} values for *Cancer pagurus* in the Irish Sea.

	G'_{ST}	D_{est}
Cpag 4	0.026	0.025
Cpag 1B9	0.000	0.000
Cpag 1C8	0.004	0.000
Cpag 2A5B	0.000	0.000
Cpag 3A2	0.025	0.015
Cpag 3D7	0.019	0.014
Cpag 4C1	0.002	0.001
Cpag 5D8	0.040	0.036
Cpag 6C4B	0.000	0.000
Cpag 15	0.001	0.000
Cpag 38	0.038	0.035

Table 3.9. D_{est} pairwise comparisons for *Cancer pagurus* in the Irish Sea.

	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX
AW													
CB	0.001												
LIA	0.000	0.000											
LIB	0.000	-0.003	0.000										
LNTZ	0.001	0.001	0.001	-0.001									
NW	0.010	0.000	0.000	-0.002	0.000								
SW	0.007	0.002	0.001	0.002	0.011	0.006							
ND	0.007	0.000	0.000	-0.001	0.000	0.001	0.001						
SD	0.002	0.000	0.004	-0.001	0.000	0.014	0.002	-0.001					
WF	0.009	0.000	0.002	-0.001	0.008	0.000	0.012	0.000	0.002				
WEX	0.002	0.000	0.000	-0.006	0.000	0.000	0.003	0.000	0.003	0.000			
M	0.001	0.000	0.001	-0.004	0.002	0.001	0.006	0.000	-0.001	0.000	0.002		
OX	0.006	0.002	0.004	0.000	0.000	0.000	0.002	0.000	-0.002	0.000	0.001	0.001	

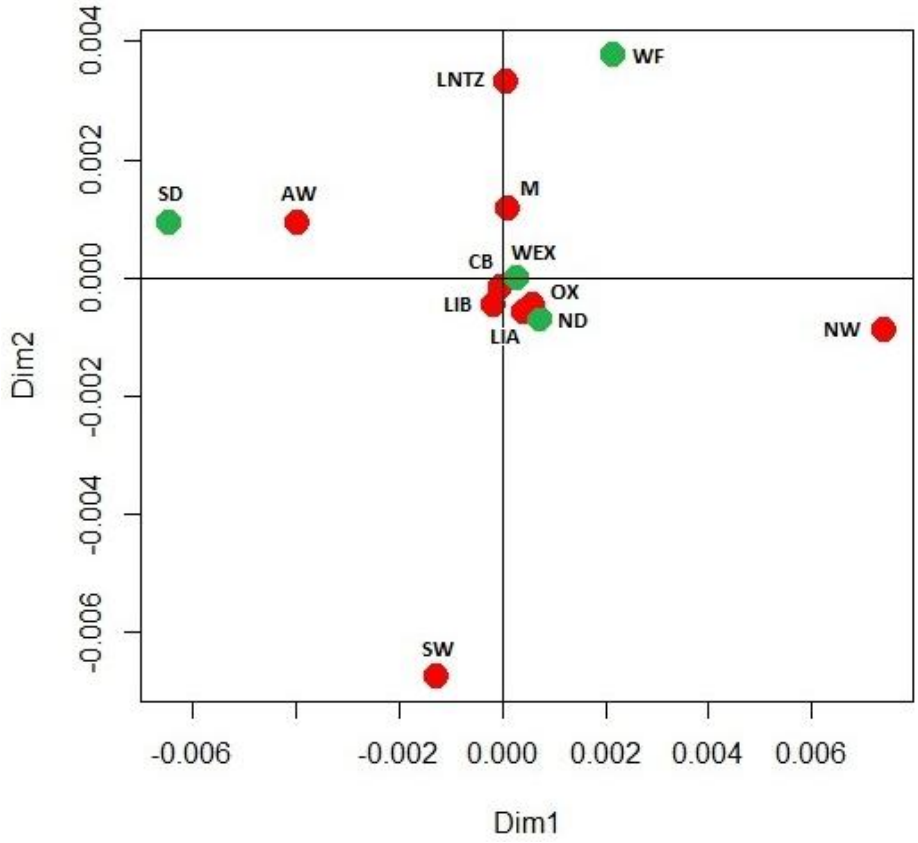


Figure 3.9. MDS plot of pairwise D_{est} values of *Cancer pagurus* in the Irish Sea. Red circles are eastern Irish Sea localities and green denotes western Irish Sea localities.

The results from the AMOVA indicate no spatial genetic structure between Wales and Ireland, as > 99.8% of the variation was found within populations (Table 3.10).

Table 3.10. AMOVA results of the Wales vs. Ireland spatial grouping.

Source of variation	Sum of squares	Variance components	Percentage of variation
Among groups	3.451	-0.00105	-0.03108
Among populations within groups	42.507	0.00658	0.19577
Within populations	3383.499	3.35787	99.83531
Total	3429.457	3.3634	100

The assessment of the relationship between the genetic distance matrix (F_{ST}) and the corresponding matrix of geographic distances (Appendix 7), revealed no evidence of IBD in the Irish Sea (r = 0.005, P = 0.48) (Figure 3.10).

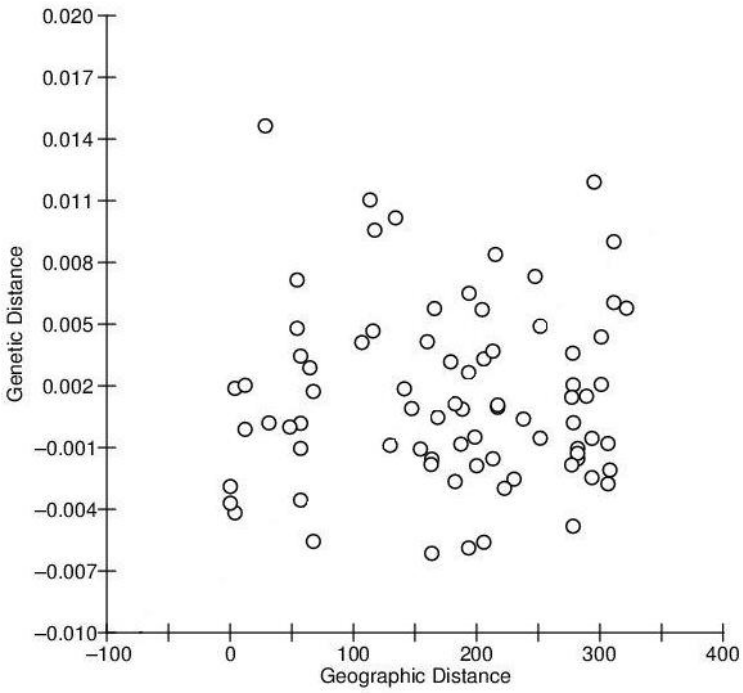


Figure 3.10. IBD in *Cancer pagurus* in the Irish Sea, with genetic distance (F_{ST}) and geographic distance (km).

Estimates of effective population size suggest that the sampled populations were very large, as all of the confidence intervals included infinity (Table 3.11). The population with the lowest N_e was CB with 443.7 (95% CI: 148.1 - ∞). N_e was large when the samples were pooled (2130; 95% CI: 1119.2 – 11865.6). BOTTLENECK detected WEX (P = 0.042) as a population with potential excess heterozygosity, using the IAM model (Table 3.12). However, the allele frequency distribution for all populations was the normal L-shape. The *M*-ratio values were below the threshold for all populations, signifying that populations might have undergone a recent population reduction (M_{AW} = 0.14, M_{CB} = 0.21, M_{LIA} = 0.23, M_{LIB} = 0.19, M_{LNTZ} = 0.20, M_{NW} = 0.22, M_{SW} = 0.18, M_{ND} = 0.22, M_{SD} = 0.23, M_{WF} = 0.22, M_{WEX} = 0.20, M_{M} = 0.19, M_{OX} = 0.24).

Table 3.11. Estimates of effective population size (N_e) in *Cancer pagurus* in the Irish Sea, with the 95% confidence intervals (parametric). The lowest allele frequency used was 0.02.

	N_e	95% CI
AW	-141.5	62.0 - ∞
CB	443.7	148.1 - ∞
LIA	791.7	182.3 - ∞
LIB	1387.7	87.2 - ∞
LNTZ	-560.8	148.3 - ∞
NW	-4127.8	241.5 - ∞
SW	-99.4	166.4 - ∞
ND	-707.6	410.7 - ∞
SD	-614.5	443.0 - ∞
WF	681.6	153.8 - ∞
WEX	-701.7	345.5 - ∞
M	-8976.8	224.5 - ∞
OX	-1946.5	298.3 - ∞
TOTAL	2130.0	1119.2 - 11865.6

Table 3.12. Assessing heterozygosity excess in *Cancer pagurus* in the Irish Sea using a Wilcoxon test under the Infinite Allele Model (IAM), the Stepwise Mutation Model (SMM) and the Two-Phase Mutation Model (TPM) (90%, variance 10). Significant values are marked with an asterisk ($P < 0.05$).

	IAM	SMM	TPM
AW	0.116	0.688	0.539
CB	0.120	0.897	0.768
LIA	0.260	0.913	0.768
LIB	0.260	0.913	0.880
LNTZ	0.103	0.861	0.861
NW	0.207	0.992	0.949
SW	0.087	0.897	0.768
ND	0.232	0.966	0.913
SD	0.483	0.994	0.913
WF	0.483	0.999	0.994
WEX	**0.042***	0.949	0.840
M	0.080	0.984	0.784
OX	0.289	0.994	0.973

There was no evidence for differences between male and female *C. pagurus* in the Irish Sea in any of the tested diversity parameters (Table 3.13). However, comparisons of Lundy NTZ against all other sites revealed significant differences for H_o (P = 0.040), F_{IS} (P = 0.031) and Relc (P = 0.031) (Table 3.14).

Table 3.13. Testing for differences between male and female *C. pagurus* across all sampling locations in the Irish Sea. A_R allelic richness, H_o observed heterozygosity, H_S gene diversity, F_{IS} inbreeding coefficient, F_{ST}, Rel relatedness and Relc corrected relatedness.

	A_R	H_O	H_S	F_{IS}	F_{ST}	Rel	Relc
Males	1.610	0.572	0.614	0.069	0.001	0.003	-0.148
Females	1.594	0.584	0.604	0.034	0.004	0.008	-0.071
P-value	0.363	0.248	0.310	0.092	0.325	0.307	0.094

Table 3.14. Comparing Lundy NTZ against all other sites within the Irish Sea. A_R allelic richness, H_o observed heterozygosity, H_S gene diversity, F_{IS} inbreeding coefficient, F_{ST}, Rel relatedness and Relc corrected relatedness.

	A_R	H_O	H_S	F_{IS}	F_{ST}	Rel	Relc
Lundy NTZ	5.926	0.533	0.617	0.135	-0.007	-0.013	-0.313
Other Populations	6.146	0.581	0.611	0.049	0.001	0.001	-0.103
P-value	0.336	0.040*	0.614	0.031*	0.137	0.170	0.031*

3.3.2. Genetic Diversity in the Northeast Atlantic

H_e ranged from 0.625 in the AW population to 0.688 in the LIA population and H_o ranged from 0.577 for the LNTZ population to 0.668 for the WEX population. Harwich (6Har) was found to have the highest N_A across loci (13.38), whereas AW had the lowest (6.13). Similarly, AW had the lowest A_R value with 6.13, whereas SW had the highest with 7.04. Five populations were shown to have a significant departure from HWE: CB (F_{IS} = 0.058; 95% CI: 0.002 – 0.086), LIA (F_{IS} = 0.133; 95% CI: 0.051 – 0.204), LNTZ (F_{IS} = 0.134; 95% CI: 0.032 – 0.205), NSW (F_{IS} = 0.089; 95% CI: 0.016 – 0.145) and Pen (F_{IS} = 0.085; 95% CI: 0.035 – 0.125) (Table 3.15). Loci Cpag-4 (11 out of 24 populations), Cpag-3A2 (6 out of 24 populations) and Cpag-5D8 (5 out of 24 populations) exhibited the most deviations from HWE ($P < 0.05$). The remaining loci did not exhibit deviations from HWE in more than 3 populations (Table 3.16). Two locus-pairings were found to be in disequilibrium: Cpag-4 & Cpag-5D8; and Cpag-1B9 & Cpag-5D8 ($P < 0.05$) (Appendix 8). F_{IS} was recalculated without locus Cpag-4; three populations remained significant (LIA, LNTZ and Pen) but CB and NSW became non-significant. The removal of Cpag 5D8, the locus involved in two linkages, resulted in the F_{IS} of four populations remaining significant (LIA, LNTZ, NSW, Pen) and CB becoming non-significant (F_{IS} = 0.062; 95% CI: -0.004 - 0.101). Furthermore, removal of Cpag-5D8 did not change patterns of significance for pairwise tests of differentiation.

FreeNA did not reveal any loci as having a large chance ($r \geq 0.20$) of harbouring null alleles, though there was some evidence of a moderate likelihood for them ($0.05 \leq r < 0.20$) (Table 3.17).

The neutrality of the eight microsatellite markers was supported by the outlier tests results, with none being identified as candidates for balancing or positive selection (Figure 3.11).

Table 3.15. Genetic diversity parameters inferred from microsatellite markers for *Cancer pagurus* in the NE Atlantic. H_e and H_o expected and observed heterozygosities, N_A number of alleles, A_R allelic richness, and F_{IS} inbreeding coefficient. Significant values are marked with an asterisk (95% CI).

	H_e	H_o	N_A	A_R	F_{IS}
AW	0.625	0.600	6.13	6.13	0.042
CB	0.680	0.641	9.13	6.67	0.058*
LIA	0.688	0.598	10.00	6.97	0.133*
LIB	0.650	0.625	7.75	6.69	0.039
LNTZ	0.665	0.577	8.00	6.20	0.134*
NW	0.685	0.637	9.25	6.47	0.071
SW	0.666	0.656	7.25	7.04	0.014
ND	0.668	0.651	9.63	6.81	0.025
SD	0.641	0.628	9.88	6.72	0.020
WF	0.663	0.629	9.13	6.16	0.051
WEX	0.684	0.668	8.63	6.41	0.022
M	0.658	0.648	8.75	6.36	0.015
OX	0.665	0.638	9.63	6.67	0.041
NNW	0.657	0.601	8.00	6.37	0.086
NSE	0.646	0.624	8.38	6.46	0.034
NSW	0.646	0.589	8.63	6.17	0.089*
NAB	0.630	0.622	10.00	6.57	0.014
Pen	0.652	0.596	10.75	6.30	0.085*
7Brt	0.646	0.621	11.38	6.64	0.040
6Jer	0.642	0.653	11.00	6.58	-0.018
6Has	0.645	0.628	12.38	6.68	0.027
6Har	0.656	0.644	13.38	6.71	0.019
6Sea	0.647	0.612	10.63	6.64	0.054
NNO	0.650	0.647	11.75	6.78	0.005

Table 3.16. Hardy-Weinberg probability test for *Cancer pagurus* in the NE Atlantic. Significant values are marked with an asterisk (P < 0.05).

PHWE	Cpag 4	Cpag 1B9	Cpag 2A5B	Cpag 3A2	Cpag 3D7	Cpag 5D8	Cpag 6C4B	Cpag 15
AW	0.108	0.703	1.000	0.731	0.726	0.344	0.340	1.000
CB	**0.019***	0.343	0.777	**0.013***	0.280	0.194	0.166	0.688
LIA	**0.000***	0.667	0.298	0.828	**0.003***	**0.000***	0.733	0.279
LIB	1.000	0.566	0.052	0.700	0.665	0.115	0.942	0.726
LNTZ	**0.013***	0.231	0.669	0.080	0.978	0.329	0.083	0.262
NW	**0.041***	0.685	0.980	0.729	0.878	0.072	0.182	0.617
SW	1.000	**0.023***	0.616	0.871	0.601	0.232	0.312	0.204
ND	0.807	0.828	0.881	0.866	0.133	0.855	0.793	0.894
SD	0.153	0.497	0.571	0.409	0.550	0.515	0.897	1.000
WF	**0.010***	0.464	0.294	0.386	0.716	0.533	0.703	0.735
WEX	**0.009***	0.421	0.753	0.817	0.393	0.982	0.993	0.753
M	0.737	0.087	0.844	0.876	0.553	0.558	0.090	0.435
OX	0.230	0.889	0.092	0.836	0.682	**0.017***	0.288	**0.035***
NNW	0.070	0.231	0.287	**0.001***	0.095	0.342	0.067	**0.020***
NSE	**0.015***	0.158	0.786	0.161	0.129	0.813	**0.013***	0.179
NSW	**0.000***	0.992	1.000	1.000	0.504	**0.000***	0.207	0.830
NAB	**0.046***	0.209	0.071	**0.016***	0.529	0.314	0.117	0.311
Pen	**0.001***	0.067	0.543	**0.001***	**0.008***	0.080	0.998	0.765
7Brt	0.235	0.307	0.103	0.058	**0.039***	0.570	0.885	0.121
6Jer	0.129	0.360	0.369	0.848	0.975	0.719	**0.002***	0.580
6Has	0.476	0.875	0.685	**0.003***	0.886	0.198	0.056	0.978
6Har	0.068	0.214	0.201	**0.000***	0.701	0.110	0.678	0.189
6Sea	**0.000***	**0.022***	0.864	0.825	0.817	**0.045***	0.200	0.733
NNO	0.086	0.090	**0.005**	1.000	0.228	**0.045***	0.802	0.701

Table 3.17. The estimate of null allele frequency per locus per population of *Cancer pagurus* in the NE Atlantic using FreeNA (Bootstrap = 1000). The values in white squares have a negligible chance of containing null alleles ($r < 0.05$) and those in grey squares have a moderate chance ($0.05 \leq r < 0.20$).

	Sampling Locations											
Loci	**AW**	**CB**	**LIA**	**LIB**	**LNTZ**	**NW**	**SW**	**ND**	**SD**	**WF**	**WEX**	**M**
Cpag 4	0.022	0.000	0.046	0.000	0.049	0.058	0.000	0.000	0.017	0.057	0.047	0.000
Cpag 1B9	0.000	0.007	0.000	0.000	0.063	0.036	0.095	0.000	0.000	0.000	0.017	0.049
Cpag 2A5B	0.017	0.003	0.057	0.128	0.042	0.000	0.000	0.000	0.025	0.000	0.000	0.017
Cpag 3A2	0.000	0.069	0.031	0.043	0.119	0.000	0.017	0.000	0.000	0.000	0.000	0.000
Cpag 3D7	0.011	0.056	0.111	0.000	0.034	0.000	0.000	0.066	0.039	0.033	0.080	0.016
Cpag 5D8	0.024	0.017	0.091	0.045	0.034	0.041	0.059	0.007	0.033	0.004	0.000	0.007
Cpag 6C4B	0.000	0.013	0.032	0.000	0.000	0.026	0.000	0.000	0.000	0.045	0.000	0.000
Cpag 15	0.000	0.000	0.017	0.000	0.003	0.047	0.000	0.000	0.000	0.000	0.000	0.000

	Sampling Locations											
Loci	**OX**	**NNW**	**NSE**	**NSW**	**NAB**	**Pen**	**7Brt**	**6Jer**	**6Has**	**6Har**	**6Sea**	**NNO**
Cpag 4	0.007	0.007	0.000	0.104	0.032	0.016	0.004	0.023	0.012	0.025	0.043	0.042
Cpag 1B9	0.009	0.000	0.000	0.000	0.000	0.000	0.037	0.000	0.000	0.000	0.008	0.000
Cpag 2A5B	0.000	0.067	0.027	0.001	0.000	0.036	0.075	0.000	0.000	0.000	0.000	0.011
Cpag 3A2	0.000	0.054	0.077	0.000	0.074	0.069	0.047	0.000	0.045	0.001	0.004	0.000
Cpag 3D7	0.000	0.000	0.102	0.046	0.017	0.087	0.000	0.000	0.016	0.000	0.000	0.012
Cpag 5D8	0.091	0.030	0.023	0.032	0.024	0.029	0.001	0.000	0.010	0.014	0.015	0.012
Cpag 6C4B	0.000	0.102	0.000	0.035	0.033	0.001	0.001	0.000	0.011	0.013	0.045	0.000
Cpag 15	0.000	0.071	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.010	0.001	0.000

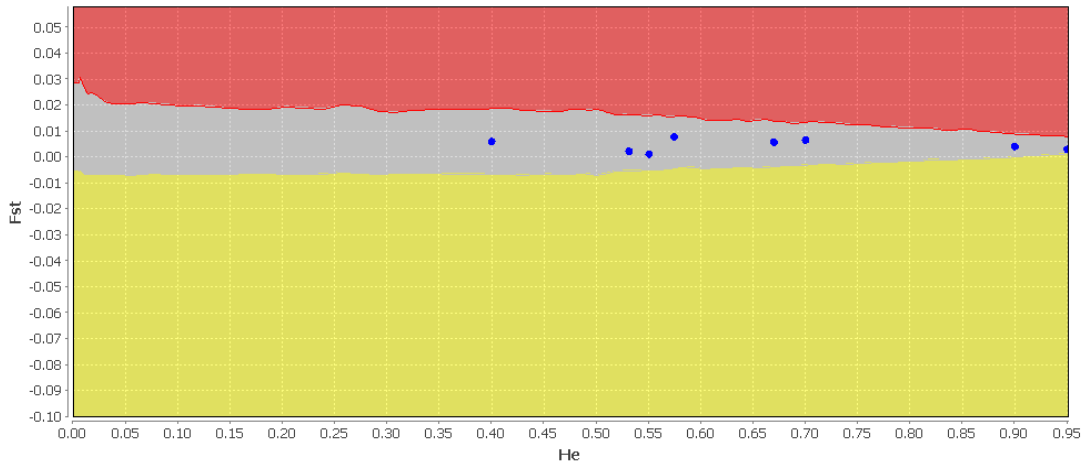


Figure 3.11. LOSITAN neutrality test for the eight microsatellite makers (blue dots) used on *Cancer pagurus* in the NE Atlantic. Markers are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area).

A mean likelihood of K = 1 (LnP(K) = -34151.62) (Figure 3.12) and an equal assignment of each individual to K clusters (Figure 3.13) gives no support of any population structuring.

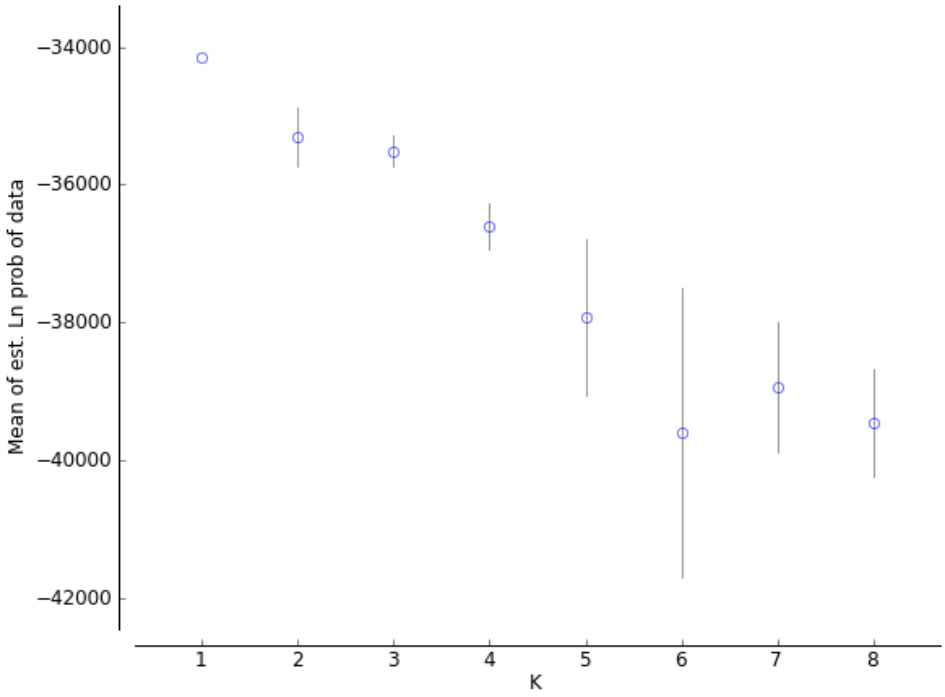


Figure 3.12. Plot of mean likelihood L(K) and variance per K value from STRUCTURE on the expanded study of the NE Atlantic containing 1433 individuals genotyped for 8 microsatellite loci.

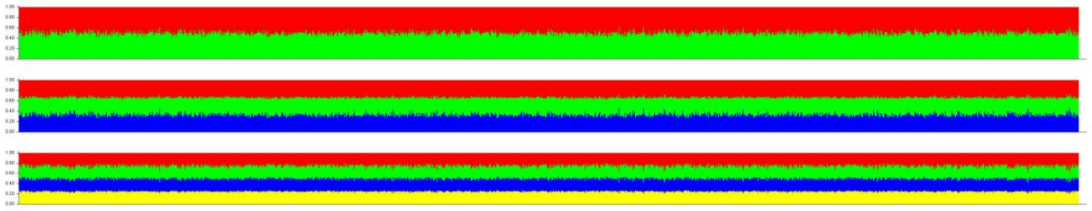


Figure 3.13. STRUCTURE bar plots for *Cancer pagurus* in the NE Atlantic for K = 2, K = 3 and K = 4 for 1433 individuals showing assignment to K clusters. Each vertical bar represents an individual and each colour a different cluster.

Global F_{ST} was significant at 0.003 (95% CI: 0.002 – 0.005). However, only 5 significant pairwise comparisons, out of a possible 276, were revealed after sequential Bonferroni correction (initial threshold P = 0.000181) (Table 3.18): NNW and LIA; NSW and LIA; NSW and M; Pen and NNW; and Pen and NSW.

Global F_{ST} was also low and significant after ENA correction, 0.003515 (95% CI: 0.002153 - 0.005125). Although, five pairwise population comparisons became significant after ENA correction: CB and 6Jer; CB and 6Har; LIA and 6Has; SW and NAB; and SD and NNW (Appendix 9). The MDS plot of the ENA-corrected F_{ST} values revealed a similar pattern to the non-corrected F_{ST} values (ENA corrected F_{ST} – Figure 3.14, F_{ST} – Figure 3.15).

Locus-specific G'_{ST} and D_{est} values both highlighted Cpag-4 as the most genetically differentiated locus, though with only < 5.8% of the maximum differentiation possible (G'_{ST} = 0.058, D_{est} = 0.055). The remaining loci revealed low levels of differentiation, with Cpag-1B9 being the least differentiated in both measures (G'_{ST} = 0.003, D_{est} = 0.002) (Table 3.19). Pairwise D_{est} revealed low levels of differentiation that were similar to those shown by F_{ST} (Table 3.20). The MDS plots for pairwise F_{ST} (Figure 3.15) and D_{est} (Figure 3.16) were visually similar and a mantel test revealed a strong positive correlation between the two genetic diversity indices (r = 0.811, P < 0.001).

Table 3.18. F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Cancer pagurus* in the NE Atlantic. Geographical regions have been colour-coded: the Irish Sea (AW, CB, LIA, LIB, LNTZ, NW, SW, ND, SD, WF, WEX, M, OX, NSE, NAB) is red, the west coast of Ireland (NNW, NSW) is green, the English Channel & Celtic Sea (Pen, 7Brt, 6Jer, 6Has, 6Har) is blue, the North Sea (6Sea) is grey, and the Norwegian Sea (NNO) is yellow. Here, the indicative adjusted nominal level (5%) for multiple comparisons has been calculated as 0.000181 (Bonferroni correction). Significant values at the P < 0.05 level are marked in bold. Values remaining significant after Bonferroni correction are marked in bold with an asterisk.

	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX	NSE	NAB	NNW	NSW	Pen	7Brt	6Jer	6Has	6Har	6Sea	NNO
AW		0.216	0.826	0.515	0.156	0.136	0.357	0.550	0.417	0.293	0.173	0.151	0.292	0.355	0.198	**0.009**	**0.002**	0.247	0.448	0.274	0.598	0.385	0.250	0.088
CB	0.006		0.178	0.656	0.093	0.292	**0.049**	0.390	0.445	0.318	0.098	0.426	0.161	**0.009**	**0.020**	**0.002**	**0.004**	**0.011**	**0.038**	**0.020**	**0.049**	**0.045**	**0.017**	**0.018**
LIA	-0.003	0.003		0.764	**0.042**	0.618	0.454	0.769	0.328	0.599	0.420	0.097	**0.032**	0.144	**0.025**	**0.000***	**0.000***	**0.037**	**0.026**	**0.002**	0.059	**0.041**	**0.017**	**0.005**
LIB	-0.003	-0.004	-0.001		0.845	0.786	0.498	0.889	0.943	0.925	0.822	0.753	0.716	0.854	0.954	0.688	0.295	0.493	0.930	0.913	0.938	0.987	0.901	0.968
LNTZ	0.017	0.006	**0.004**	-0.004		0.593	0.130	0.319	0.302	0.222	0.222	0.076	0.335	**0.006**	**0.040**	**0.025**	**0.006**	**0.039**	0.128	**0.045**	0.299	0.573	0.448	0.116
NW	0.011	0.000	0.001	-0.005	-0.001		0.226	0.805	**0.047**	0.582	0.166	0.181	0.687	0.103	**0.047**	**0.006**	**0.005**	0.054	**0.037**	**0.003**	0.232	**0.020**	0.087	**0.035**
SW	0.005	**0.000**	-0.003	0.001	0.008	0.004		0.570	0.564	0.070	**0.018**	0.086	0.592	0.165	**0.050**	0.070	**0.028**	0.086	0.432	0.404	0.458	0.306	0.572	0.292
ND	0.010	0.005	-0.002	-0.001	0.002	0.003	-0.004		0.947	0.586	0.382	0.231	0.858	0.306	0.091	**0.017**	**0.009**	0.138	0.572	0.476	0.993	0.745	0.530	0.581
SD	0.002	0.006	0.005	0.001	0.008	**0.013**	-0.001	0.001		0.626	0.103	0.631	0.577	**0.036**	0.164	**0.017**	**0.003**	0.490	0.530	0.431	0.895	0.620	0.453	0.432
WF	0.003	-0.001	0.002	-0.006	0.006	0.002	0.002	0.000	0.001		0.671	0.300	0.567	**0.028**	0.111	0.078	**0.008**	0.634	0.137	0.243	0.374	0.490	0.136	0.506
WEX	0.006	-0.001	-0.001	-0.005	-0.001	-0.001	**-0.002**	0.001	0.005	-0.001		0.076	0.051	**0.018**	0.080	**0.021**	**0.022**	**0.026**	0.187	**0.013**	0.164	0.199	**0.002**	**0.049**
M	0.005	-0.003	0.002	-0.005	0.003	0.002	-0.002	-0.002	-0.002	-0.003	0.001		0.168	**0.027**	0.057	**0.002**	**0.000***	0.328	0.400	0.062	0.149	0.500	0.156	0.631
OX	0.003	0.001	**0.003**	-0.003	0.005	0.003	-0.007	-0.001	-0.001	-0.002	0.001	-0.001		0.295	**0.009**	**0.038**	**0.021**	0.089	0.599	0.651	0.236	0.131	0.445	0.136
NSE	0.005	**0.004**	0.008	0.001	**0.023**	0.004	0.003	0.007	**0.016**	**0.005**	**0.009**	**0.007**	0.005		0.097	0.462	**0.002**	**0.000***	0.098	**0.012**	**0.002**	0.066	**0.005**	**0.005**
NAB	0.004	**0.009**	**0.011**	-0.007	**0.014**	**0.009**	**0.012**	0.008	0.007	0.003	0.008	0.005	**0.008**	0.009		**0.007**	**0.001**	**0.009**	0.197	**0.041**	**0.017**	0.106	0.113	**0.038**
NNW	**0.014**	**0.010**	**0.014***	0.002	**0.019**	**0.007**	0.002	**0.008**	**0.010**	0.002	**0.007**	**0.009**	**0.002**	0.003	**0.009**		0.122	**0.000***	**0.001**	**0.005**	**0.009**	**0.002**	**0.000**	**0.009**
NSW	**0.029**	**0.007**	**0.023***	0.004	**0.016**	**0.012**	**0.010**	**0.016**	**0.018**	**0.008**	**0.008**	**0.012***	**0.008**	**0.019**	**0.019**	0.003		**0.012**	0.058	**0.015**	0.086	0.082	0.088	0.181
Pen	-0.001	**0.006**	**0.000**	-0.002	**0.006**	0.008	0.000	0.000	-0.001	0.000	**0.003**	-0.001	0.001	**0.013***	**0.010**	**0.024***	**0.005**		**0.010**	**0.008**	0.083	0.478	0.149	0.274
7Brt	-0.001	**0.003**	**0.006**	-0.004	0.011	**0.005**	0.001	0.003	0.000	0.000	0.003	0.000	-0.001	0.003	0.002	**0.014**	0.003	**0.002**		0.649	0.181	0.442	**0.014**	0.151
6Jer	0.007	**0.006**	**0.006**	-0.002	**0.006**	**0.008**	0.001	0.001	0.003	0.001	**0.003**	0.002	0.000	**0.008**	**0.012**	**0.009**	**0.009**	**0.004**	0.003		**0.033**	**0.040**	0.064	**0.005**
6Has	0.003	**0.005**	0.005	-0.003	0.008	0.008	-0.002	0.000	-0.002	0.000	0.003	-0.001	-0.001	**0.006**	**0.010**	**0.012**	0.003	0.001	0.000	**0.001**		0.545	0.184	0.268
6Har	0.006	**0.004**	**0.005**	-0.005	0.007	**0.007**	0.000	0.001	0.001	0.000	0.003	-0.001	0.000	0.005	0.008	**0.011**	0.003	0.001	0.002	**0.003**	0.000		0.347	0.597
6Sea	0.004	**0.005**	**0.004**	-0.005	0.002	0.007	-0.001	-0.001	-0.001	0.001	**0.005**	-0.002	-0.001	**0.011**	0.011	**0.016**	0.003	-0.002	**0.002**	0.001	-0.002	-0.002		0.751
NNO	0.016	**0.006**	**0.011**	0.000	0.009	**0.011**	0.003	0.000	0.002	0.001	**0.007**	-0.002	0.002	**0.010**	**0.012**	**0.014**	0.005	0.002	0.003	**0.006**	0.001	0.000	-0.002	

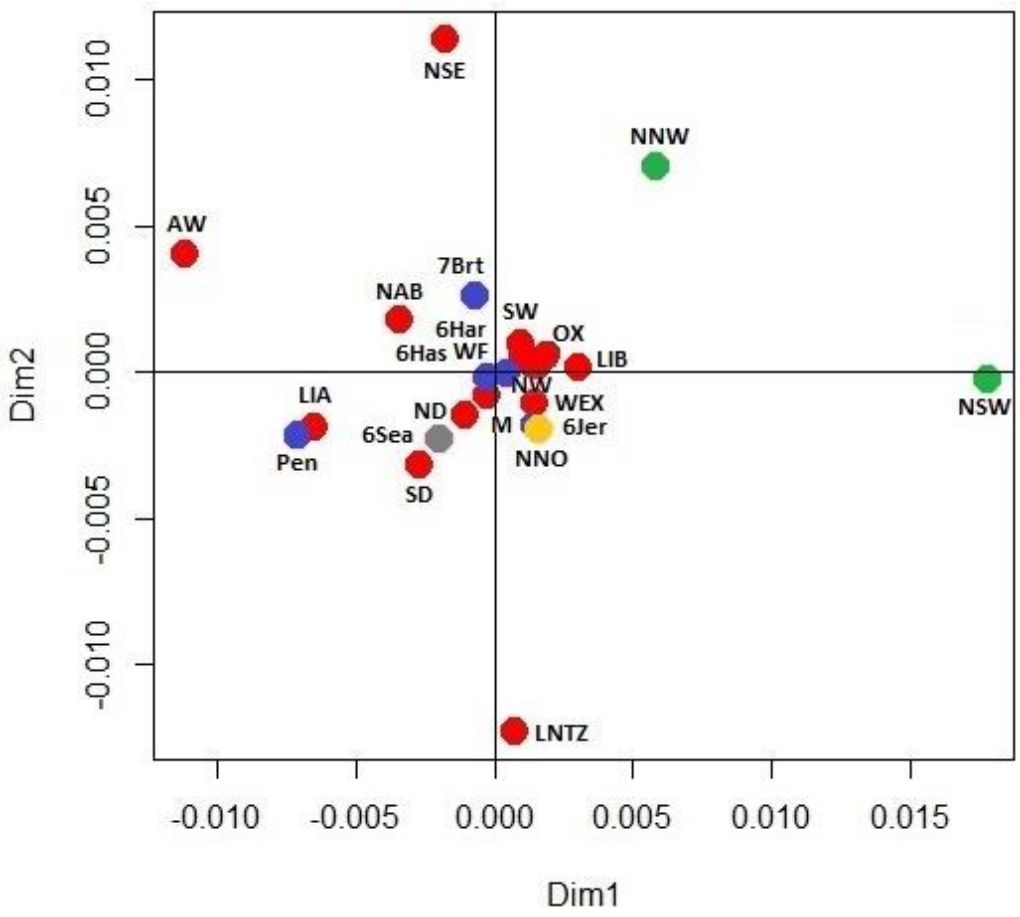


Figure 3.14. MDS plot of ENA-corrected pairwise F_{st} values of *Cancer pagurus* in the NE Atlantic. Geographical regions have been colour-coded: the Irish Sea (AW, CB, LIA, LIB, LNTZ, NW, SW, M, OX, ND, SD, WF, WEX, NSE, NAB) is red, the west coast of Ireland (NNW, NSW) is green, the English Channel (Pen, 7Brt, 6Jer, 6Has, 6Har) is blue, the North Sea (6Sea) is grey, and the Norwegian Sea (NNO) is yellow.

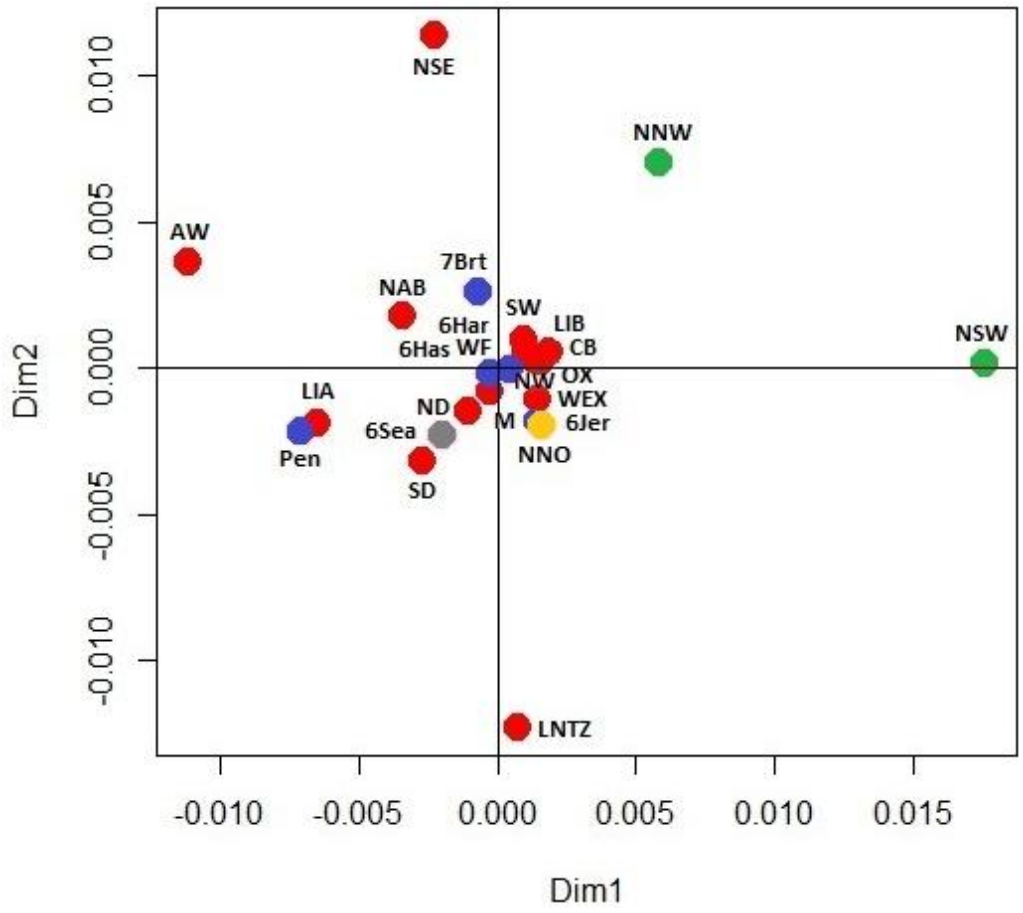


Figure 3.15. MDS plot of pairwise F_{st} values of *Cancer pagurus* in the NE Atlantic. Geographical regions have been colour-coded: the Irish Sea (AW, CB, LIA, LIB, LNTZ, NW, SW, M, OX, ND, SD, WF, WEX, NSE, NAB) is red, the west coast of Ireland (NNW, NSW) is green, the English Channel (Pen, 7Brt, 6Jer, 6Has, 6Har) is blue, the North Sea (6Sea) is grey, and the Norwegian Sea (NNO) is yellow.

Table 3.19. Locus-specific G'_{ST} and D_{est} values for *Cancer pagurus* in the NE Atlantic.

	G'_{ST}	D_{est}
Cpag 4	0.058	0.055
Cpag 1B9	0.003	0.002
Cpag 2A5B	0.005	0.003
Cpag 3A2	0.018	0.010
Cpag 3D7	0.021	0.015
Cpag 5D8	0.041	0.037
Cpag 6C4B	0.017	0.012
Cpag 15	0.010	0.004

Table 3.20. D_{est} pairwise comparisons for *Cancer pagurus* in the NE Atlantic.

	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX	NNW	NSE	NSW	NAB	Pen	7Brt	6Jer	6Has	6Har	6Sea	NNO
AW																								
CB	0.000																							
LIA	0.000	0.000																						
LIB	0.000	-0.001	0.002																					
LNTZ	0.002	0.001	0.002	-0.003																				
NW	0.006	0.001	0.000	0.000	0.000																			
SW	0.003	0.001	0.000	0.003	0.011	0.005																		
ND	0.004	0.001	0.000	-0.001	0.000	0.002	0.000																	
SD	0.002	0.000	0.005	0.000	0.000	0.021	0.000	-0.002																
WF	0.004	0.001	0.005	0.000	0.007	0.003	0.009	0.000	0.003															
WEX	0.001	0.000	0.001	-0.005	0.000	0.001	0.003	0.000	0.004	0.000														
M	0.000	0.000	0.003	-0.003	0.001	0.004	0.002	0.000	-0.002	0.000	0.001													
OX	0.005	0.002	0.007	0.000	0.000	0.001	0.000	-0.002	-0.002	0.000	0.002	0.001												
NNW	0.023	0.017	0.027	0.000	0.025	0.012	0.010	0.006	0.013	0.002	0.011	0.009	0.002											
NSE	0.000	0.004	0.007	0.000	0.015	0.001	0.007	0.001	0.008	0.004	0.008	0.005	0.002	0.001										
NSW	0.041	0.011	0.039	0.009	0.025	0.018	0.015	0.009	0.023	0.006	0.012	0.011	0.008	0.002	0.016									
NAB	0.001	0.004	0.010	-0.008	0.002	0.004	0.014	0.002	0.002	0.003	0.000	0.001	0.005	0.013	0.006	0.024								
Pen	0.001	0.004	0.003	0.000	0.004	0.009	0.003	0.000	0.000	0.001	0.003	-0.001	0.003	0.013	0.005	0.021	0.005							
7Brt	0.000	0.002	0.011	-0.002	0.003	0.003	0.002	0.000	0.000	0.001	0.001	0.000	0.000	0.003	0.000	0.015	0.001	0.002						
6Jer	0.004	0.004	0.009	-0.001	0.001	0.011	0.004	-0.001	0.000	0.000	0.001	0.000	0.000	0.003	0.003	0.004	0.002	0.003	0.000					
6Has	0.000	0.001	0.005	-0.002	0.000	0.003	0.000	-0.002	-0.003	0.001	0.000	0.000	0.000	0.006	0.004	0.012	0.000	0.000	0.000	0.001				
6Har	0.003	0.004	0.007	-0.004	0.001	0.009	0.003	0.000	0.000	0.001	0.005	0.000	0.002	0.005	0.000	0.009	0.000	0.000	0.000	0.000	0.000			
6Sea	0.006	0.003	0.006	-0.005	0.000	0.004	-0.001	-0.001	0.000	0.003	0.006	0.000	0.000	0.011	0.002	0.013	0.001	0.000	0.002	0.001	-0.001	-0.001		
NNO	0.014	0.007	0.014	-0.001	0.003	0.011	0.005	0.000	0.000	0.000	0.002	0.000	0.000	0.008	0.008	0.008	0.001	0.000	0.000	0.000	0.000	0.000	-0.001	

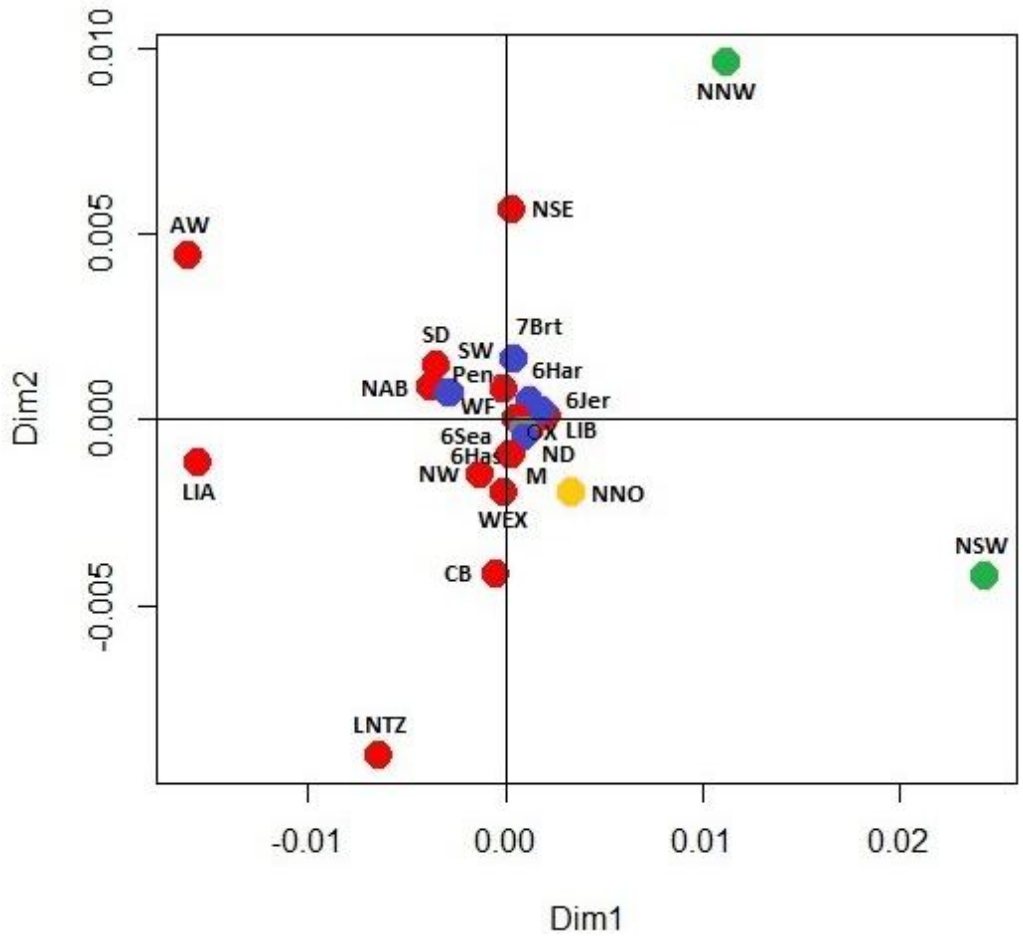


Figure 3.16. MDS plot of pairwise D_{est} values of *Cancer pagurus* in the NE Atlantic. Geographical regions have been colour-coded: the Irish Sea (AW, CB, LIA, LIB, LNTZ, NW, SW, M, OX, ND, SD, WF, WEX, NSE, NAB) is red, the west coast of Ireland (NNW, NSW) is green, the English Channel (Pen, 7Brt, 6Jer, 6Has, 6Har) is blue, the North Sea (6Sea) is grey, and the Norwegian Sea (NNO) is yellow.

No evidence of IBD was detected (r = 0.155, P = 0.15) (Figure 3.17, Appendix 10).

Estimates of effective population size indicate that most of the sampled populations were very large, on account of many of the confidence intervals including infinity (Table 3.21). However, NSW was shown to have a small and finite effective size of 45.7 (95% CI: 30.1 – 80.1). BOTTLENECK detected LNTZ (P = 0.037) and WEX (P = 0.020) as populations with potential heterozygosity excess using the IAM model (Table 3.22). Though, the allele frequency distribution for all populations was the normal L-shape. No significant heterozygosity excess was detected using the SMM or TPM. The *M*-ratio values were below the assumed threshold for all populations, signifying a population reduction (M_{AW} = 0.11, M_{CB} = 0.16, M_{LIA} = 0.18, M_{LIB} = 0.14, M_{LNTZ} = 0.15, M_{NW} = 0.17, M_{SW} = 0.13, M_{ND} = 0.16, M_{SD} = 0.17, M_{WF} = 0.16, M_{WEX} = 0.15, M_{M} = 0.15, M_{OX} = 0.17, M_{NNW} = 0.14, M_{NSE} = 0.14, M_{NSW} = 0.16, M_{NAB} = 0.18, M_{Pen} = 0.19, M_{7Brt} = 0.19, M_{6Jer} = 0.19, M_{6Has} = 0.22 M_{6Har} = 0.24, M_{6Sea} = 0.19, M_{NNO} = 0.20).

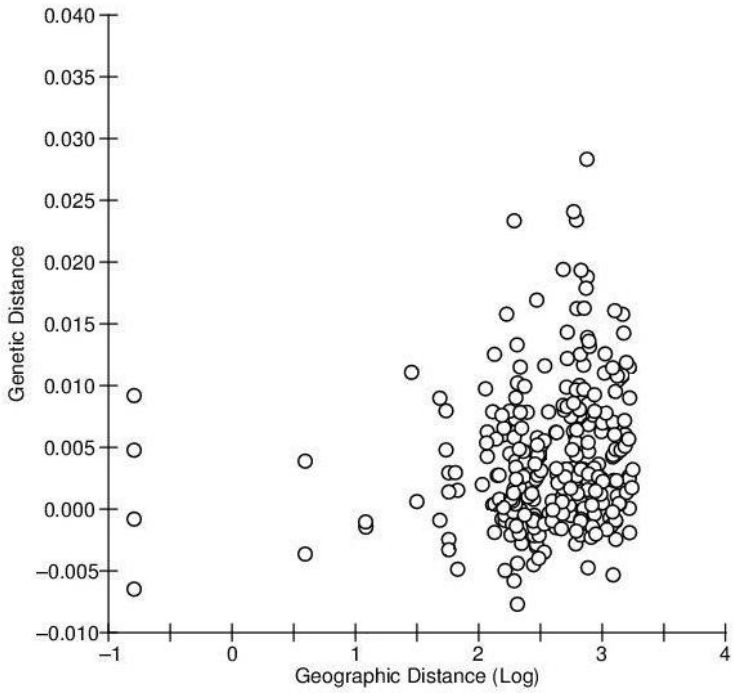


Figure 3.17. IBD in *Cancer pagurus* in the NE Atlantic, with genetic distance (F_{ST}) and log geographic distance (km).

Table 3.21. Estimates of effective population size (N_e) in *Cancer pagurus* in the NE Atlantic, with the 95% confidence intervals (parametric). The lowest allele frequency used was 0.02.

	N_e	95% CI
AW	-95.5	49.6 - ∞
CB	222.2	92.9 - ∞
LIA	597.4	131.2 - ∞
LIB	1164.6	60.5 - ∞
LNTZ	-377.1	108.2 - ∞
NW	-489.7	261.3 - ∞
SW	-168.7	57.7 - ∞
ND	-830.0	240.3 - ∞
SD	-846.5	239.5 - ∞
WF	-5707.2	149.0 - ∞
WEX	-355.9	318.6 - ∞
M	-258.8	704.0 - ∞
OX	-1261.0	220.1 - ∞
NNW	-1050.2	78.1 - ∞
NSE	210.2	47.7 - ∞
NSW	45.7	30.1 - 80.1
NAB	-377.2	408.6 - ∞
Pen	2594.2	295.5 - ∞
7Brt	-766.5	820.9 - ∞
6Jer	-727.4	412.6 - ∞
6Has	-1375.0	796.0 - ∞
6Har	-1632.5	725.0 - ∞
6Sea	-223140.4	186.4 - ∞
NNO	-303.1	1550.7 - ∞
TOTAL	-20663.3	6148.7 - ∞

Table 3.22. Assessing heterozygosity excess in *Cancer pagurus* in the NE Atlantic using a Wilcoxon test under the Infinite Allele Model (IAM), the Stepwise Mutation Model (SMM) and the Two-Phase Mutation Model (TPM) (90%, variance 10). Significant values are marked with an asterisk ($P < 0.05$).

	IAM	SMM	TPM
AW	0.191	0.527	0.422
CB	0.125	0.527	0.422
LIA	0.273	0.809	0.527
LIB	0.125	0.680	0.680
LNTZ	**0.037***	0.680	0.680
NW	0.156	0.963	0.902
SW	0.125	0.809	0.680
ND	0.191	0.770	0.680
SD	0.422	0.963	0.809
WF	0.273	0.990	0.963
WEX	**0.020***	0.875	0.875
M	0.098	0.902	0.578
OX	0.191	0.980	0.963
NNW	0.125	0.902	0.875
NSE	0.371	0.727	0.680
NSW	0.125	0.994	0.986
NAB	0.273	0.980	0.902
Pen	0.371	0.986	0.902
7Brt	0.371	0.973	0.902
6Jer	0.371	0.994	0.963
6Has	0.422	0.994	0.973
6Har	0.422	0.998	0.994
6Sea	0.371	0.980	0.875
NNO	0.527	0.994	0.980

3.4. DISCUSSION

Simulation analysis indicated that the microsatellite data (loci and sample sizes) conferred sufficient statistical power to detect even very low levels of true genetic differentiation. Within the Irish Sea, variation at 11 microsatellite loci revealed no significant population differentiation, compatible with high gene flow throughout the region. On a wider geographic scale, Bayesian clustering analysis based on eight microsatellite loci supported a model of one genetic population, indicating a high level of background gene flow throughout the studied region.

Despite the overall pattern of high gene flow the expanded study did report a number of significant tests that fitted with a pattern of 'chaotic genetic patchiness' (Johnson & Black 1982), with populations in close geographic proximity differing genetically by as much as those hundreds of kilometres apart. The biological interpretation of such low levels of genetic differentiation revealed by microsatellite loci is notoriously difficult for marine taxa (Waples 1998; Knutsen et al. 2011). As statistical power is high when combining information from multiple highly polymorphic markers and large sample sizes, minor frequency differences that are not biologically meaningful can achieve statistical significance (Ryman & Palm 2006), even though POWSIM analysis indicated a low probability of Type I errors. Furthermore, when true differentiation is low or absent various confounding factors may assume greater significance and contribute to erroneous inferences of local populations (Waples 1998). These factors may be technical, such as genotype scoring errors (Bonin et al. 2004), or involve non-random sampling of individuals, e.g. sampling family or kin aggregations (Allendorf & Phelps 1981). However, such technical artefacts can be considered highly unlikely here. Firstly, the microsatellite markers provided unambiguous genotypes that were confirmed by at least two operators. Integration of the Irish Sea and NE Atlantic data sets was carried out by analysis of 96 control specimens to ensure compatible genotype calling. Furthermore, with the exception of Cpag-4, the microsatellite loci provided limited evidence of null alleles that may generate false positives (Shaw et al. 2010), with patterns persisting upon correction for null alleles and exclusion of Cpag-4.

Secondly, as samples consisted of mixed cohorts of adults the probability of sampling related individuals must be considered low.

Chaotic genetic patchiness has been reported in other marine species with planktonic larvae and in some cases has been attributed to temporal variation in the genetic composition of recruits (Johnson & Black 1982; Hedgecock 1994; Knutsen et al. 2003; Pujolar et al. 2006; Kennington et al. 2013). Such variation among/within recruits may stem from large variances in reproductive success, i.e. sweepstakes reproductive success (Hedgecock 1994). Many highly fecund marine taxa exhibit large variances in reproductive success which could generate fine-scale patchy genetic structuring, even in the absence of restricted dispersal (Johnson et al. 1993). In the case of *C. pagurus*, McKeown & Shaw (2008b) suggest that female sperm usage patterns and fishing practices may further increase variance in reproductive success. Paternity analysis of broods from 18 ovigerous females reported no evidence of multiple paternity, despite the capacity of the species for long-term storage of sperm and suspected potential for females to use sperm from multiple males simultaneously (McKeown & Shaw 2008b). Within such a genetically monogamous system the number of females places a strict constraint on the number of males that can breed. In the *C. pagurus* fishery, female landings per unit effort are considerably higher than for males for most of the year (Bennett 1995), which may further increase variance in reproductive success.

In addition to variation in reproductive success, other nonmutually exclusive hypotheses to explain chaotic genetic patchiness include (i) variation in the source of larval recruits, (ii) pre-and/or post settlement natural selection, and (iii) population isolation and self-recruitment (Toonen & Grosberg 2011). Hogan et al. (2010) and Selkoe et al. (2006) both reported chaotic patchiness in marine taxa arising from variability in the sources of larval recruits. However, in light of the overall weak genetic structure reported for *C. pagurus*, it is unlikely that sources of larval recruits will be genetically distinct populations, but rather that genetic differences among groups will have been generated by sweepstakes-like reproductive skews. Spatial genetic patchiness could be due to selection which has been shown to be a potentially powerful agent in the genetic structuring of marine

populations (Conover et al. 2006). The wide distribution of *C. pagurus* is likely to encompass highly varied environmental conditions providing the opportunity for local adaptation to develop; several studies have found evidence for local adaptation in widely distributed marine fish (Nielsen et al. 2009). However, outlier analysis provided no evidence of selection effects on the employed loci, though genome scans, candidate gene analysis or population transcriptomics would be needed to more robustly test this possible explanation. Finally, it cannot be ruled out that the chaotic pattern reflects population isolation (i.e. breakdown in gene flow) occurring within a non-migration-drift equilibrium system wherein patterns of structuring may be determined by differences in effective population sizes, demographic history, migration, or some combination of these factors, and poorly reflect contemporary connectivity patterns (Lowe & Allendorf 2010).

In this respect, the prominence of the west of Ireland samples among significant pairwise tests, despite the overriding pattern of high connectivity throughout the sampled region, is interesting. Specifically, (1) Donegal and Lundy Island; (2) Galway Bay and Lundy Island; (3) Galway Bay and Mumbles; (4) Pendeen and Donegal; and (5) Pendeen and Galway Bay, were significant. This may reflect some restricted gene flow between the west of Ireland and the Celtic Sea, which could be the result of a barrier to adult migration. Sotelo et al. (2008) reported the significant differentiation of a sample from the west of Ireland from more southern samples. Female *C. pagurus* are known to migrate against the current to enable the planktotrophic larvae to hatch in prevailing tidal currents, thus facilitating the return of settling larvae to areas of maternal origin (Pawson 1995; Eaton et al. 2003). The Irish coastal current along the western coast of Ireland is a continuous pathway existing during the summer months from north Cornwall to Malin Head, the most northerly point of Ireland (Fernand et al. 2006). Female *C. pagurus* from the west of Ireland would migrate against this current, with their larvae ultimately returning northwards once released. Also, migrations from the Celtic Sea to western Ireland would not be occurring, as movement would be with the current. Therefore, these factors could result in the observed pattern. Larval dispersal models and tagging studies would be helpful in verifying these hypotheses about adult

movements, as prior studies have focused on the English Channel (Bennett & Brown 1983; Hunter et al. 2013).

It is frequently shown for the N_e of marine populations to be two to five orders of magnitude smaller than the census population size (N_c), which explains how collapsed fish stocks could lose genetic diversity regardless of the large spawning stock biomasses (Hauser et al. 2002). Results generated with the BOTTLENECK and M-ratio analyses differed; Wexford and Lundy NTZ were the only sampling locations shown to have heterozygosity excess, whereas the M-ratio values were below the threshold for all populations. The tests in BOTTLENECK revealed heterozygosity excess only under an IAM model, which is considered an unlikely mutational model for microsatellites (Piry et al. 1999). Also, these samples did not reveal reduced N_e estimates. It is recommended in order to be statistically conservative to use only SMM when analysing microsatellite data to test for recent bottlenecks (Luikart & Cornuet 1998) and so the results for Wexford and Lundy NTZ must be considered with caution. An important consideration in the interpretation of the low M-ratio values is that the critical value of M was assumed and not calculated specifically.

Despite the fact that shellfisheries are increasing, the numbers of shellfish are also increasing due to fishing practices removing higher trophic level predators (Molfese et al. 2014). Therefore, it is important to know the N_e of *C. pagurus* and to manage the resource effectively to prevent the possibility of stock collapse. Estimates of effective population size were very large for the most part, with all confidence intervals including infinity, apart from Galway Bay in the expanded study which had a finite size of 45.7. The small N_e detected at Galway Bay is of concern, since it is generally considered that a N_e of 50 individuals is the minimum required for short-term conservation of heterozygosity, with 500 individuals needed for more long-term considerations of adaptability (Hauser et al. 2002). However, as this sample did not exhibit any signatures of chronic genetic erosion (e.g. lower levels of polymorphism, or significant bottleneck tests) it may be that the low N_e estimate actually reflects a low N_b (i.e. effective number of breeders) due to occurrence of sweepstakes recruitment wherein few individuals (breeders) successfully contribute recruits. Coastal topography has been shown to influence recruitment patterns,

even in long dispersing taxa (Banks et al. 2007). As such, sweepstake recruitment might be expected to be more pronounced for the Galway sample as this was the only sample collected within a semi-enclosed bay. Consequently, this may explain the prominent differentiation and low N_e estimate of this sample.

Despite only moderate levels of null alleles recorded across loci, heterozygote deficiency ($H_o < H_e$) was observed across sampling locations. Furthermore, positive significant F_{IS} at two locations in the Irish Sea (Lundy Island and Lundy Island NTZ) and, with the removal of Cpag-4, three locations in the NE Atlantic expanded study (Lundy Island, Lundy Island NTZ, and Pendeen). Such deficits are compatible with the aforementioned variances in reproductive success. In light of the results from Chapter 2, which reported elevated F_{IS} at the Lundy NTZ in *H. gammarus*, it is interesting that this site also revealed significantly positive F_{IS} values among *C. pagurus*. The signatures of increased variance in a NTZ may be explained by the persistence of larger females who produce more offspring (Tallack 2007a) and thus greater potential for skews of allele frequencies. Migrant analysis (BAYESASS) revealed insufficient power to accurately quantify migration between samples (data not shown). However, under the assumption of a high proportion of allochthonous larval recruits entering the NTZ, the increased variance in reproductive success may be due to more patchy post-settlement survival associated with density dependent factors (discussed in Chapter 2).

Determining the population structure of *C. pagurus* with both a fine-scale microsatellite analysis of the Irish Sea, plus a broad expanded study in the Northeast Atlantic has indicated a high degree of connectivity throughout the studied range. This is compatible with findings for the holoplanktonic species *Sagitta setosa* (Peijnenburg et al. 2006), a number of marine invertebrates with high larval dispersal potential (Triantafyllidis et al. 2005; Couceiro et al. 2007; Sotelo et al. 2008; Sotelo et al. 2009), as well as flatfish (Hoarau et al. 2002; Hoarau et al. 2004). Against this background of high connectivity there was some evidence for chaotic genetic patchiness. Resolution of the specific drivers of this pattern will require analysis of temporal sampling and ideally age partitioned cohorts. The implications of such patchiness for management are discussed by Larson & Julian

(1999). The indication that this pattern is driven by large variances in reproductive success implies that recruitment may be patchy over time. Therefore, management should endeavour to ensure that the distribution and potential of the total spawning stock is protected. The implied spatial stochasticity also has relevance for the effectiveness of marine reserves and would suggest they be spatially dispersed. However, the results for the Lundy NTZ may indicate that such reserves introduce additional stochasticity and a better approach would be to ensure suitably low exploitation throughout the species range. Finally, the detected patchiness may significantly underestimate the extent of recruitment variability due to the sampling of adults or, indicate that postlarval ontogenetic dispersal is much lower than previously thought. In both cases, this highlights the susceptibility of *C. pagurus* populations/stocks to environmental or anthropogenic impacts despite significant gene flow.

4. PRELIMINARY INVESTIGATION OF EDIBLE CRAB, *CANCER PAGURUS*, RADSEQ VARIATION IN THE NORTHEAST ATLANTIC

4.1. INTRODUCTION

Many marine species, including some of the most economically important taxa, exhibit features such as high dispersal potential, high fecundity, large population sizes and wide geographic distributions. These 'classical' (Nielsen & Kenchington 2001) marine traits, in conjunction with the lack of conspicuous barriers to dispersal in the marine environment (Briggs 1974), traditionally fostered expectations of high gene flow, sparse local adaptation and low levels of genetic drift culminating in limited opportunities for population divergence (Palumbi 1994; Ward et al. 1994; Waples 1998). However, population genetic studies of marine taxa are providing increased evidence of significant intraspecific neutral and adaptive structuring. Such structuring may be driven by mechanisms that are extrinsic (e.g. oceanographic fronts) or intrinsic to a species, or an interplay of both, and be significant to varying degrees, on ecological and evolutionary time frames. Determination of the extent of such structuring and its contributing factors are fundamental to the ability to effectively manage and conserve stocks and their ecosystems (Palumbi 2003), and are also essential to predicting responses to harvesting and environmental changes. Such information is urgently needed for marine taxa with many such species subjected to increased harvesting pressures and exhibiting spatial/temporal demographic changes linked to ecosystem fishery effects (Molfese et al. 2014) and climate change (MacKenzie et al. 2014).

The edible crab, *Cancer pagurus*, is distributed continuously in shallow shelf waters of the NE Atlantic from the Lofoten Islands (northern Norway) to Morocco (Bennett 1995). The species supports one of the most important commercial fisheries in Northern European waters and its socio-economic importance has contributed to extensive research on its biology, movements and fisheries (reviewed in Bennett 1995). Tagging of adults has revealed marked differences between male and female dispersal; males are largely resident, making short seemingly random movements

within small territories while females exhibit much more complex patterns of dispersal. Females are reported to undertake very long migrations (hundreds of kilometres)(Hunter et al. 2013), with information across the species distribution indicating that these movements are predominantly directed against the prevailing currents in the respective areas (Pawson 1995; Eaton et al. 2003). This feature has also been observed among artificially transplanted individuals (Bennett 1995; Ungfors et al. 2007). These migrations are seemingly one way with limited evidence for return movements (but see Ungfors et al. 2007 for discussion on possible shortcomings of tagging data). Females are highly fecund with estimates of 0.5 to 2.9 million eggs being produced per female per brood, with larger females producing more eggs (Edwards 1979; Ungfors 2007). Moreover, their capacity for sperm storage enables a single supply of sperm to be used to fertilise multiple batches of eggs over a number of years (McKeown & Shaw 2008b). Hatched larvae are pelagic for 1-3 months, depending on water temperature (Eaton et al. 2003).

C. pagurus thus, exhibits a number of characteristics expected to limit population structuring throughout its range. Concordant with this expectation, no significant genetic differentiation was reported among samples spanning 1300 km of waterway distance from the Norwegian Sea, Skagerrak and Kattegat (Ungfors et al. 2009). Furthermore, Chapter 3 of this thesis also reported non-significant genetic differentiation among samples from the Irish Sea. However, population structuring can vary throughout a species range (Cowen et al. 2006). Larval surveys and oceanographic modelling of crab larvae have identified a frontal system in the North Sea, known as the Flamborough front, as a barrier to larval dispersal (Eaton et al. 2003). Marked spatial heterogeneities in larval abundances reported in the English Channel (D. Eaton, unpublished) and North Sea (Eaton et al. 2003) may also indicate some non-panmictic recruitment dynamics. Furthermore, while connectivity may be affected by adult (female) dispersal, little is known about the influence of the contranatant migrations on subsequent larval recruitment (Bennett 1995).

Population genetic studies at broad and local geographical scales are needed for *C. pagurus*. As a preliminary investigation of such broader scale patterns, Chapter 3,

though focused on the Irish Sea, included microsatellite data for samples collected from throughout the NE Atlantic. The microsatellite data indicated that while the majority of pairwise tests were non-significant, signifying a high level of gene flow, there were a number of significant pairwise tests contributing to a pattern of 'chaotic genetic patchiness' (Johnson & Black 1982). The results of Chapter 3 epitomises the difficulties in interpreting the biological significance of low levels of genetic differentiation among marine species (Knutsen et al. 2011). In particular for microsatellites, as statistical power is high when employing multiple highly polymorphic loci and large sample sizes (Ryman & Palm 2006), minor allele frequency differences that are not associated with population structure, and thus not 'biologically meaningful', can achieve statistical significance (Waples 1998).

Next Generation Sequencing (NGS) has provided unprecedented access to genomic resources (Mardis 2008) but remains a costly method for marker development in non-model organisms. Furthermore, for population genetic based studies whole genome sequencing is not required, with genotyping of loci distributed throughout the genome favoured. As such there has been considerable interest in the utility of reduced representation genome sequencing methods such as Restriction-site Associated DNA sequencing (RADseq) for genetic analysis of non-model taxa. However, despite the potential and increasing popularity of RADseq (Rowe et al. 2011) it has largely been limited to model, or emerging model taxa (Baird et al. 2008; Emerson et al. 2010; Hohenlohe et al. 2010; Hohenlohe et al. 2011) with only a single marine invertebrate RADseq data set published to date (sea anemone, *Nematostella vectensis*) (Reitzel et al. 2013).

The results from Chapter 3 indicated that, although conferring a high degree of statistical power, microsatellite data are unlikely to provide robust inferences of biologically significant differences among populations that could be readily integrated into ongoing management plans. Using loci under selection could assist with the identification of ecologically and genetically relevant units and, therefore, are considered here as a way to build on the findings from Chapter 3. An objective of this research was therefore to investigate the potential of Restriction-site Associated DNA (RAD) as a future tool for population genetic analysis and fisheries

management of *C. pagurus* and other widely distributed commercial invertebrates, with particular interest in the potential of using genome scans to: (i) provide increased resolution of neutral population structure and (ii) identify molecular signatures of divergent selection. A secondary aim was to generate *a posteriori* hypotheses of potential neutral and adaptive structuring processes.

4.2. MATERIALS AND METHODS

4.2.1. Study Area and Sample Collection

Edible crab were sampled from 13 locations across the NE Atlantic (N = 178), with sites chosen so that a large portion of the distributional range was encompassed (Table 4.1, Figure 4.1). Sampling locations were along the west coast of Ireland [Donegal (NWI) and Galway Bay (SWI)], the Irish Sea [Howth (ISA) and Lundy Island NTZ (ISB)], the Celtic Sea (NQ), the English Channel [Brittany (Brit), Hastings (Hast) and Harwich (Har)], Shetland (Shet), the Norwegian Sea [(LL), (LV), (G)] and Gulmarsfjorden (Gul). Males were selected when possible due to the known migratory behaviour of female *C. pagurus*.

Genomic DNA was either purified from haemolymph or tissue using the QIAGEN DNeasy Blood & Tissue Kit (QIAGEN, CA-USA), or had previously been extracted through phenol-chloroform (freezer-stored samples). DNA was quantified via a Qubit fluorometer (Life Technologies) with the Broad Range (BR) kit by adding 5 µl DNA to 195 µl working solution (1:200) of 199 µl QUANT-iT dsDNA BR buffer and 1 µl BR reagent. DNA sample dilutions were then standardised to provide 100 ng per sample for downstream sequencing.

Table 4.1. The location, assigned code and coordinates of the sampling locations, as well as the date of sample collection (S) and the number of individuals sampled (N).

Location	Code	Coordinates	S	N
Donegal, NW Ireland (IE)	NWI	54.558° N, 8.315° W	Jul-07	12
Galway Bay, SW Ireland (IE)	SWI	53.196° N, 9.279° W	Jul-07	12
Howth, North Dublin, Irish Sea (IE)	ISA	53.469° N, 6.084° W	Jul-11	12
Lundy Island NTZ, Bristol Channel (UK)	ISB	51.189° N, 4.649° W	May-10	12
Pendeen, Celtic Sea (UK)	NQ	50.153° N, 5.740° W	Jun-06	12
Brittany, English Channel (FR)	Brit	47.250° N, 5.500° W	Jul-06	12
Hastings, English Channel (UK)	Hast	50.718° N, 0.662° E	Oct-06	12
Harwich, North Sea (UK)	Har	51.590° N, 1.590° E	May-05	24
Shetland (UK)	Shet	60.230° N, 0.788° W	Jun-10	24
West Norway, Norwegian Sea (NO)	LL	62.672° N, 6.663° E	Dec-04	8
Lunneviken, Skagerrak	LV	59.058° N, 11.167° E	Oct-06	8
Grove Bank, Kattegat	G	57.104° N, 11.512° E	Jul-07	8
Gullmarsfjorden (SW)	Gul	58.783° N, 11.183° E	Jul-07	22

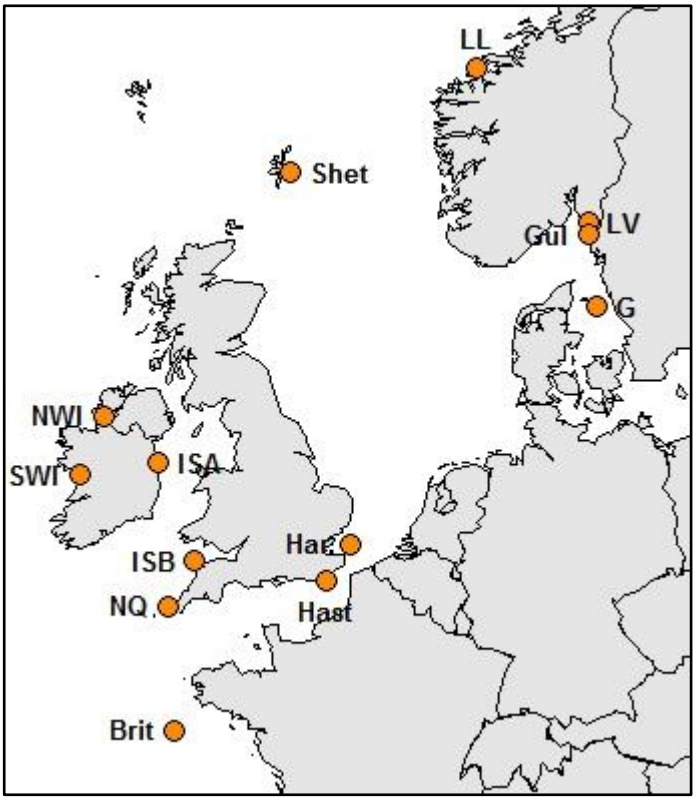


Figure 4.1. Sampling locations of the edible crab *Cancer pagurus* in the NE Atlantic.

4.2.2. Creation of RAD Tag Libraries

Library preparation was carried out using a modified version of the RADseq protocol that was first outlined by Baird et al. (2008). For each individual, genomic DNA (200 ng) was digested for 1 hour at 37 °C in a 50 µl reaction with 20 units of *Pst*I (CTGCA*G) (New England Biolabs [NEB]) and 5 µl of buffer 4 (NEB). The enzyme was heat-inactivated at 65 °C for 30 minutes. Individuals that were to be sequenced together in a particular run were individually 'barcoded' to permit downstream demultiplexing of sequence data. This was carried out by annealing P1 adaptors that included five nucleotide barcode sequences (Table 4.2). 2 µl of 25 nM P1 Adaptor (Illumina, Inc.) was added to each sample along with 1 µl of 100 mM rATP (Promega), 1 µl of buffer 2 (NEB), 0.5 µl of T4 DNA ligase (NEB), 7 µl of H_2O and incubated at room temperature for 1 hour, then left overnight at 4 °C. Samples were again heat-inactivated for 20 minutes at 65 °C, pooled, purified (QIAquick PCR Purification Kit, QIAGEN), and then sheared (Bioruptor Plus, Diagenode). A further purification (Agencourt AMPure XP, Beckman Coulter) and elution in 22 µl buffer EB

(QIAGEN) preceded the size selection with a LabChip XT DNA 750 following the Assay Quick Guide protocol (Caliper Life Sciences) to isolate DNA 250 bp ± 10% and 305 bp ± 10%. The DNA from the size select was purified and eluted in 20 µl H_2O. The Quick Blunting Kit (NEB) was used to remove overhanging sequences with a 30 minute room temperature incubation. Samples were purified and eluted in 16 µl H_2O before the addition of 2 µl Klenow exo- (NEB), 2 µl buffer 2 and 1 µl dATP (NEB), and then incubated for 30 minutes at 37 °C to add adenine overhangs on the 3' end of the DNA. After another purification and elution in 21 µl H_2O, 1 µl of 10 µM P2 Adaptor (Illumina, Inc.) was ligated to the DNA fragments at room temperature for 20 minutes, with 3 µl of buffer 2, 1 µl of 100 mM rATP, 0.5 µl of T4 DNA ligase and 4 µl H_2O. Samples were again purified and eluted in 50 µl of buffer EB. PCR amplification was carried out six times per sample (1.25 µl per reaction), each with a final volume of 25 µl, containing 12.5 µl Phusion Master Mix (NEB), 2.5 µl P1/P2 PCR Primer Mix (Illumina, Inc.) and 8.75 µl of H_2O. The PCR cycle involved an initial step at 98 °C for 30 seconds, followed by 19 cycles of 10 seconds at 98 °C, 30 seconds at 65 °C and 30 seconds at 72 °C, and a final step at 72 °C for 5 minutes, with the PCR products held at 4 °C to ensure optimal amplification. The six reactions per sample were pooled, purified and eluted in 12 µl of buffer EB. A second size selection with a LabChip XT DNA 750 isolated DNA 310 bp ± 10% and 380 bp ± 10% before a final clean-up and elution in 20 µl buffer EB.

4.2.3. Sequencing Platform and SNP Calling

Sequencing was performed on the Illumina HiScanSQ platform (now discontinued); a modified scanner with a sequencing fluidics module added on, enabling it to process HiSeq flow cells (Table 4.2). Sequencing adaptors and barcodes were removed using Floragenex (Eugene, OR, USA) software tools, resulting in 94-bp RAD fragments. Variants were called using the SAMtools (Li et al. 2009) variant detection pipeline and the subsequent pileup files were parsed using custom scripts. Custom scripts and VCF tools (Danecek et al. 2011) were used to further filter the single-nucleotide variant data based on within-stack alignment statistics and conformity of genotype frequencies to Hardy-Weinberg expectations. With data analysis focusing

on detecting patterns of population genetic structure, the resultant SNP dataset passed relatively stringent filtering criteria (Table 4.3).

Table 4.2. Barcode and lane order for each sample on the Illumina HiScanSQ platform (Lanes 1-7).

Barcode	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6
AGAGT	ISA_01	NQ_52	Brit_01	Har_09	Shet_01	G_11
AGCTG	ISA_02	NQ_53	Brit_02	Har_10	Shet_02	G_13
AGGAC	ISA_03	NQ_54	Brit_03	Har_11	Shet_03	G_14
ACTCA	ISA_04	NQ_59	Brit_04	Har_12	Shet_04	G_15
ATATC	ISA_05	NQ_60	Brit_05	Har_13	Shet_05	G_16
ATCGA	ISA_06	NQ_62	Brit_06	Har_14	Shet_06	G_18
ATGCT	ISA_07	NQ_64	Brit_07	Har_15	Shet_07	G_19
ATTAG	ISA_08	NQ_65	Brit_13	Har_16	Shet_08	G_21
CGATA	ISA_09	NQ_75	Brit_14	Har_17	Shet_09	LV_01
CGCGC	ISA_10	NQ_84	Brit_16	Har_18	Shet_10	LV_04
CGGCG	ISA_11	NQ_86	Brit_17	Har_19	Shet_11	LV_05
CGTAT	ISA_12	NQ_87	Brit_20	Har_20	Shet_12	LV_07
CTAGG	ISB_01	SWI_03	Hast_09	Har_22	Shet_13	LV_09
CTCTT	ISB_02	SWI_07	Hast_10	Har_23	Shet_14	LV_12
CTGAA	ISB_03	SWI_10	Hast_11	Har_24	Shet_15	LV_13
CTTCC	ISB_04	SWI_11	Hast_12	Har_25	Shet_16	LV_14
GGAAG	ISB_05	SWI_12	Hast_13	Har_26	Shet_17	LL_01
GGCCT	ISB_06	SWI_13	Hast_14	Har_27	Shet_18	LL_03
GGGGA	ISB_07	SWI_14	Hast_15	Har_28	Shet_19	LL_05
GGTTC	ISB_08	SWI_15	Hast_17	Har_35	Shet_20	LL_08
GTACA	ISB_09	SWI_16	Hast_19	Har_36	Shet_21	LL_29
GTCAC	ISB_10	SWI_18	Hast_21	Har_37	Shet_22	LL_31
GTGTG	ISB_11	SWI_19	Hast_22	Har_38	Shet_23	LL_33
GTTGT	ISB_12	SWI_24	Hast_26	Har_39	Shet_24	LL_35

Lane 7					
Barcode	**Sample**	**Barcode**	**Sample**	**Barcode**	**Sample**
ATGCT	NWI_10	CGGCG	Gul_1	GCGCC	Gul_13
ATTAG	NWI_12	CGTAT	Gul_2	GCTAA	Gul_14
CAACT	NWI_14	CTAGG	Gul_3	GGAAG	Gul_15
CACAG	NWI_16	CTCTT	Gul_4	GGCCT	Gul_16
CAGTC	NWI_17	CTGAA	Gul_5	GGGGA	Gul_17
CATGA	NWI_18	CTTCC	Gul_6	GGTTC	Gul_18
CCAAC	NWI_20	GAAGC	Gul_7	GTACA	Gul_19
CCCCA	NWI_24	GACTA	Gul_8	GTCAC	Gul_20
CCGGT	NWI_25	GAGAT	Gul_9	GTGTG	Gul_21
CCTTG	NWI_26	GATCG	Gul_10	GTTGT	Gul_22
CGATA	NWI_28	GCATT	Gul_11		
CGCGC	NWI_29	GCCGG	Gul_12		

Table 4.3. Filtering criteria for SNPs in *Cancer pagurus*: Q is the minimum Phred-like single-nucleotide variant score (Li et al. 2008); Min. depth is the minimum number of reads; Missing (%) is the maximum percentage of missing genotype data allowed for a given locus; Max. F_{IS} is the maximum deviation of observed genotype frequencies from Hardy-Weinberg expectations, and; Min. het. reads is the minimum proportion of reads supporting the less frequent allele in a heterozygous genotype. Table modified from Slavov et al. (2014).

Filtering Criteria	
Q	15
Min. depth	30
Missing (%)	5
Max. F_{IS}	0.1
Min. het. reads	0.05

4.2.4. Statistical Analysis

RAD marker variability was compared among individuals within each sample using five commonly estimated genetic indices: (1) the mean expected heterozygosity, (2) the mean observed heterozygosity, (3) the mean number of alleles per locus, (4) the number of polymorphic SNPs, and (5) F_{IS}. These indices were calculated using the R-package POPGENKIT v1.0 (Rioux Paquette 2011).

Two methods were employed to identify potential markers in genomic regions under selection. The F_{ST} outlier method, FDIST2, of Beaumont & Nichols (1996) was implemented in LOSITAN (Antao et al. 2008) with simulations run for 10^5 iterations, 99% CI, and performed for the neutral mean F_{ST} under the infinite alleles mutation model. The false discovery rate (FDR) was set at 0.1. Outlier detection was also performed in BAYESCAN (Foll & Gaggiotti 2008), which uses a hierarchical Bayesian approach to estimate locus and population effects on F_{ST} values. As suggested by Foll & Gaggiotti (2008), outlier loci were identified using the 'decisive' criterion under Jeffrey's scale of evidence to minimise the false-positive rate and to maximise the true-positive rate. Both outlier tests were performed globally and between pairs of samples, with a criterion of a strict consensus between both methods enforced to identify and exclude potentially non-neutral markers for further analysis. This criterion led to the formation of two datasets for independent analysis: (1) a

'neutral' dataset, with consensus markers under balancing and positive selection removed; and (2) an 'adaptive' dataset comprised of consensus outlier loci under positive selection.

Sample allele frequencies for each locus were calculated in GENEPOP 4.2.1 (Raymond & Rousset 1995; Rousset 2008). Estimations of locus-specific F_{ST} values and 95% CIs (100 bootstraps) were determined with the R-package DIVERSITY (Keenan et al. 2013) to denote SNPs that contribute most to distinguishing samples.

Global and pairwise F_{ST} values were calculated in Arlequin 3.5 (Excoffier & Lischer 2010). The significance of *F*-statistics was tested by 1000 permutations. F_{ST} values were then visualised using a Principal Coordinate Analysis (PCoA) performed in R 3.0.2 (R Core Team 2013) for the neutral dataset and each divergent outlier loci, separately.

Mantel tests, as implemented in the IBDWS software (Jensen et al. 2005), were used to test for correlation between genetic (F_{ST}) and geographical (km) distances between sample sites (i.e. isolation by distance). Geographic distances were measured by tracing the shortest route between two populations via sea using the program NETPAS 2.5 (Netpas). Genetic and geographical distances were analysed with tests based on 10,000 randomisations.

Principal Component Analysis (PCA) was performed to facilitate the identification of subsets of markers that effectively describe population differences (Paschou et al. 2007), using the R-package ADEGENET (Jombart 2008; Jombart & Ahmed 2011).

Genetic structure was investigated in the 'neutral' dataset using the Bayesian clustering analysis implemented in the program STRUCTURE (Pritchard et al. 2000; Falush et al. 2003, 2007). This was used to identify the number of clusters, *K* (from a range of 1-5), with the highest posterior probability. Analyses were replicated for both the 'loc prior' and 'no loc prior' under the 'admixture model with correlated allele frequencies'. Each MCMC run consisted of a burn in of 10^6 steps followed by 5 x 10^6 steps. Three replicates were conducted for each *K* to assess consistency. The *K* value best fitting the data set was estimated by the log probability of data [Pr(X/K].

4.3. RESULTS

A total of 566 SNPs were detected in *C. pagurus* in the NE Atlantic.

H_e ranged from 0.0908 in the Hast population to 0.1305 in the Gul population, and H_o ranged from 0.0598 for the Hast population to 0.0930 in the G population. Gul was found to have the highest N_A per loci (1.5512), whereas ISA had the lowest (1.2862). All populations were revealed to have significant F_{IS} values, ranging from 0.2625 in the G population to 0.4882 in the NWI population. G was shown to have the lowest number of polymorphic loci (260), compared to LV with the highest (408) (Table 4.4).

Table 4.4. Genetic diversity parameters inferred from SNPs for *Cancer pagurus* in the NE Atlantic. Expected (H_e) and observed (H_o) heterozygosities, the number of alleles per locus (N_A/locus), inbreeding coefficient (F_{IS}) and the number of polymorphic loci (Poly. Loci). Significant values are marked with an asterisk (95% CI).

	H_e	H_o	N_A / locus	F_{IS}	Poly. Loci
NWI	0.1300	0.0667	1.4576	0.4882*	300
SWI	0.1243	0.0766	1.4770	0.3987*	340
ISA	0.0986	0.0692	1.2862	0.3105*	381
ISB	0.1114	0.0801	1.3604	0.2945*	380
NQ	0.1271	0.0711	1.4329	0.4534*	343
Brit	0.1140	0.0790	1.3993	0.3204*	309
Hast	0.0908	0.0598	1.2898	0.3544*	355
Har	0.1211	0.0763	1.5088	0.3803*	362
Shet	0.1232	0.0780	1.5300	0.3748*	307
LL	0.1236	0.0838	1.3587	0.3376*	299
LV	0.1013	0.0670	1.2898	0.3566*	408
G	0.1238	0.0930	1.4028	0.2625*	260
Gul	0.1305	0.0715	1.5512	0.4656*	278

The global outlier test implemented in LOSITAN identified 11 loci as candidates for positive selection, with 8 SNPs remaining significant outliers after FDR correction (strong outliers) (P > 0.995) (Figure 4.2) (Appendix 11), with the number of divergent SNPs identified from pairwise comparisons ranging from zero (ISA vs. NQ; NQ vs. Brit; NQ vs. Hast) to 60 (ISB vs. Shet) (Table 4.5) (Appendix 12). Outlier detection performed globally in BAYESCAN highlighted four divergent SNPs (Figure 4.3), but none were identified from pairwise comparisons. Overall, there were three consensus outlier SNPs between both statistical methods: SNPs 148, 469 and 541. Spatial patterns of allele frequencies for each consensus outlier reveal no clear geographical pattern (SNP 148, Figure 4.4; SNP 469, Figure 4.5; SNP 541, Figure 4.6). Estimations of locus-specific F_{ST} values and 95% CIs for all 566 SNPs denote those that contribute most to distinguishing samples (Figure 4.7).

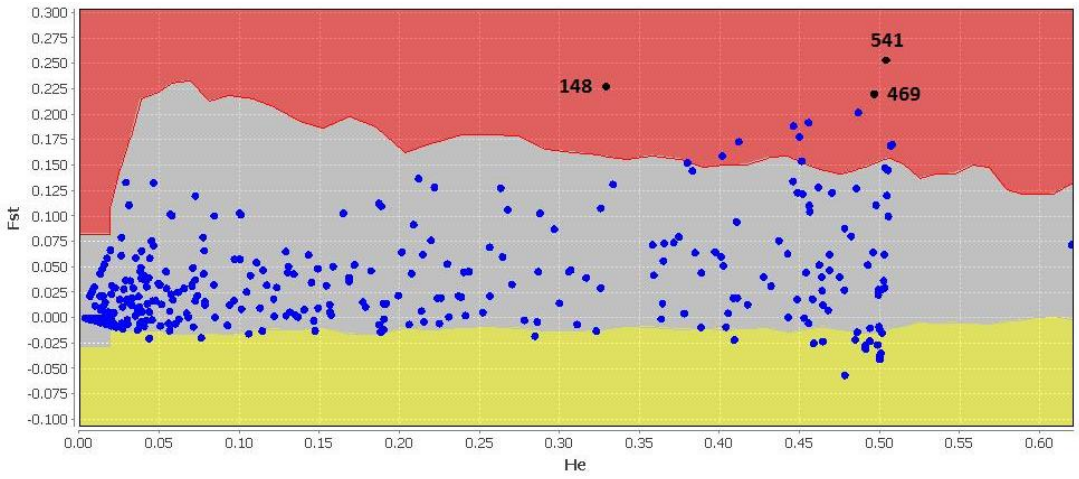


Figure 4.2. Global outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered as candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 11 SNPs were detected as positive outliers, with 8 SNPs remaining significant outliers after FDR correction. The three consensus outliers are highlighted in black.

Table 4.5. Number of divergent SNPs identified from pairwise comparisons in LOSITAN, with those remaining as outliers after FDR correction in brackets.

	NWI	SWI	ISA	ISB	NQ	Brit	Hast	Har	Shet	LL	LV	G	Gul
NWI													
SWI	7												
ISA	7 (6)	5 (4)											
ISB	21	13	19										
NQ	12	15	-	5									
Brit	19 (18)	20	11 (9)	1	-								
Hast	32	5	5	7	-	6							
Har	10	16	18 (15)	9	24	8	13						
Shet	55	15	16 (13)	60	7	8	16 (14)	4					
LL	28	39	15	21 (20)	14	23	5	16	24				
LV	27	9 (5)	3	25	33	13	14 (13)	26	26	32			
G	8	26	13	10 (5)	41	32	9 (8)	12	23	5 (4)	56		
Gul	41	22	15 (14)	16	6	12	17	6 (5)	8 (5)	37	43 (42)	25	

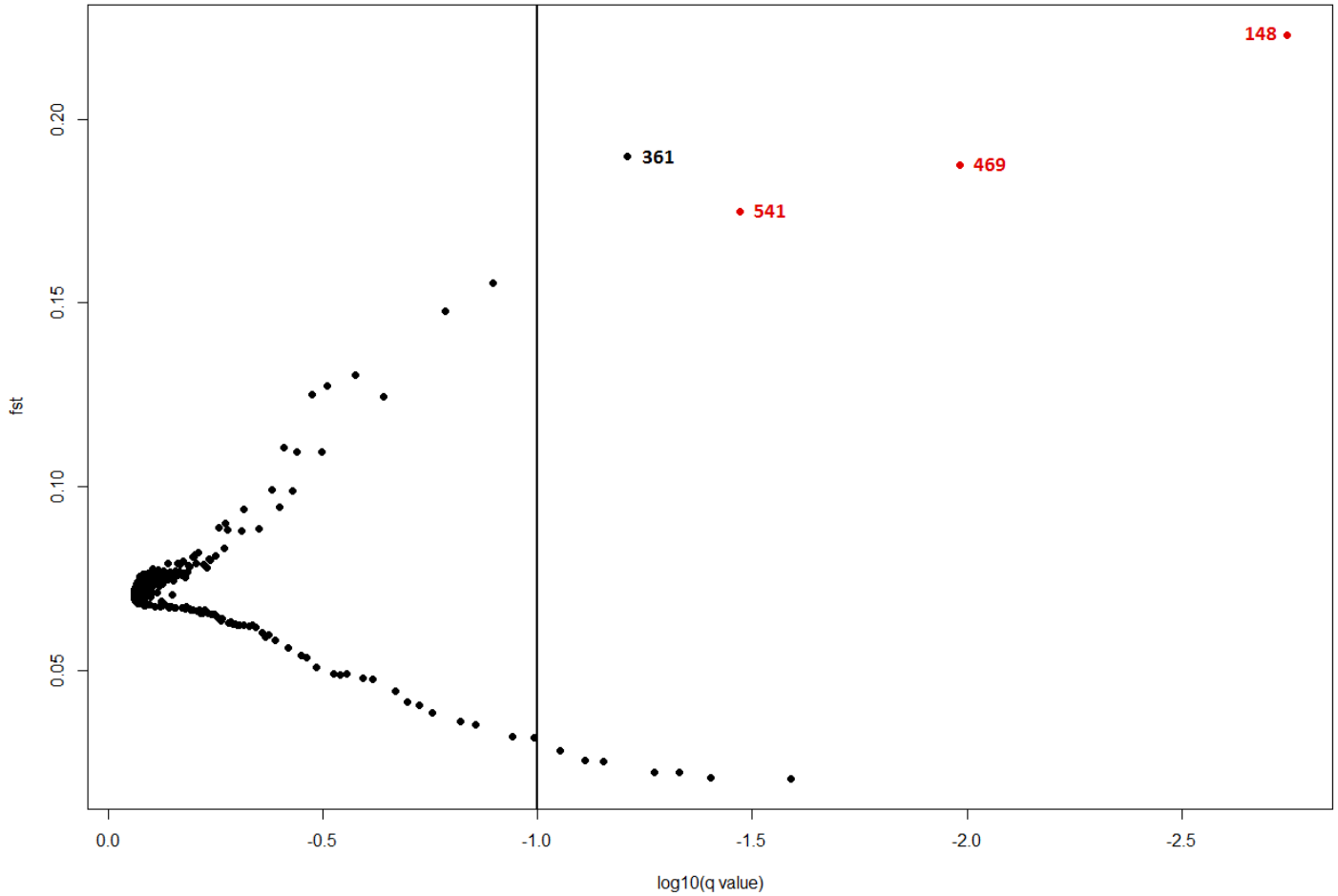


Figure 4.3. Global outlier test implemented in BAYESCAN revealing four divergent SNPs. The three consensus outliers are highlighted in red.

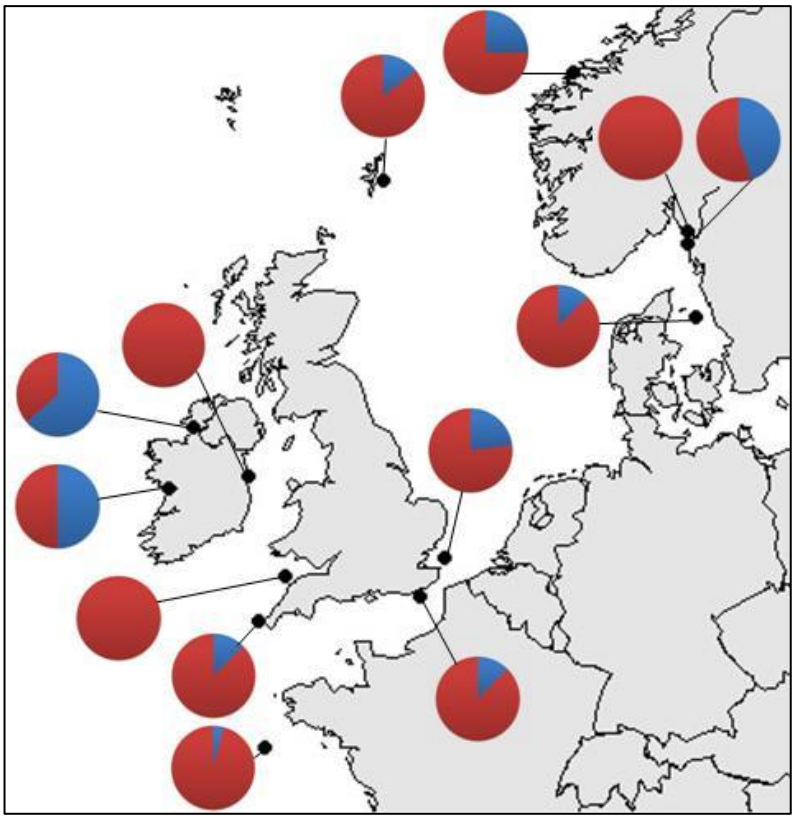


Figure 4.4. SNP 148 allele frequencies.

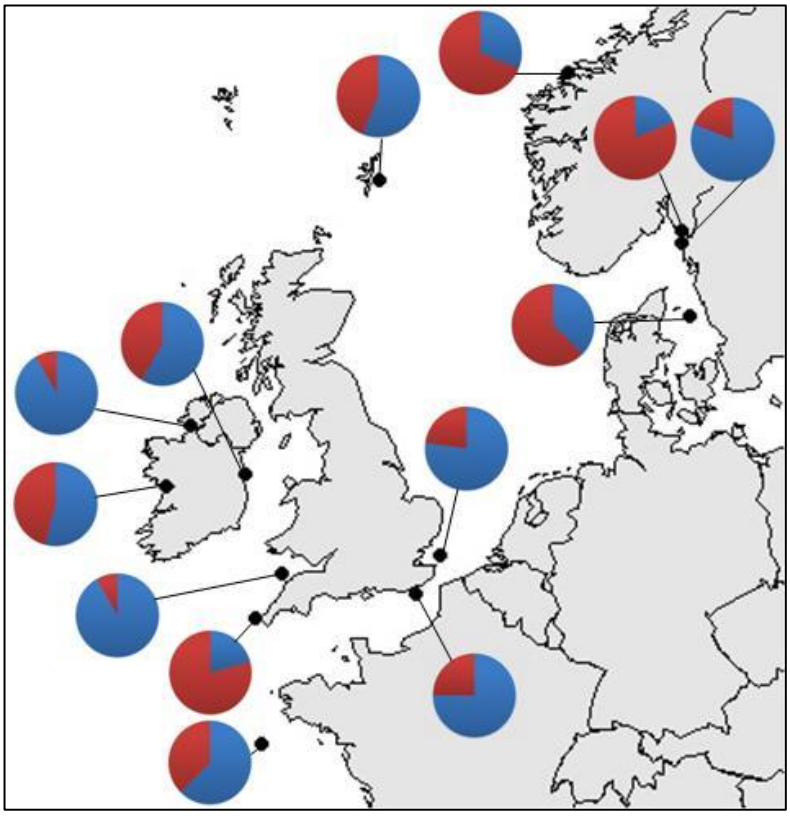


Figure 4.5. SNP 469 allele frequencies.

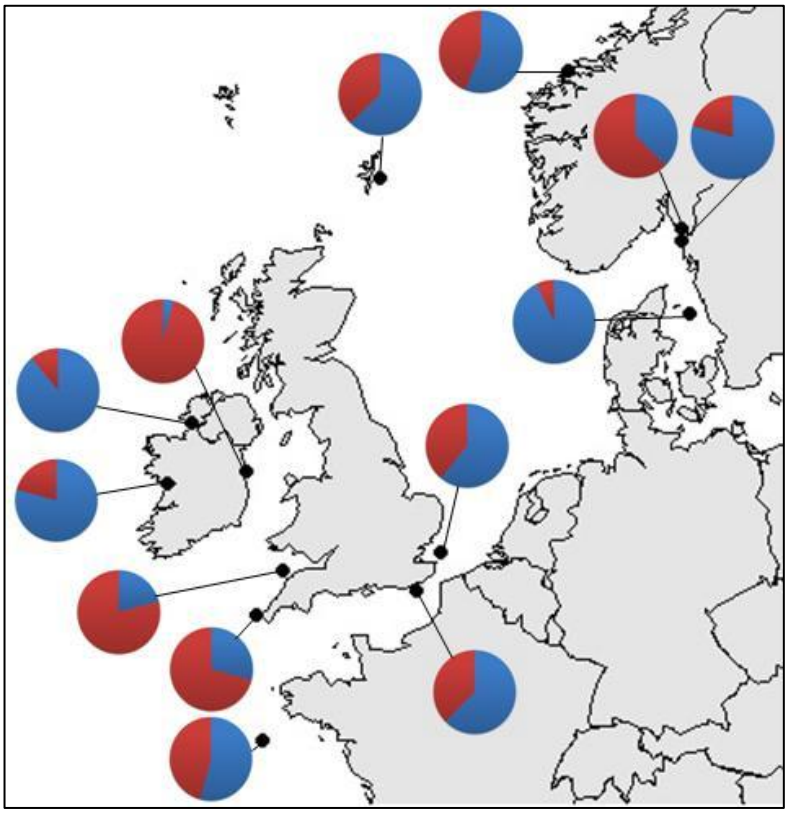


Figure 4.6. SNP 541 allele frequencies.

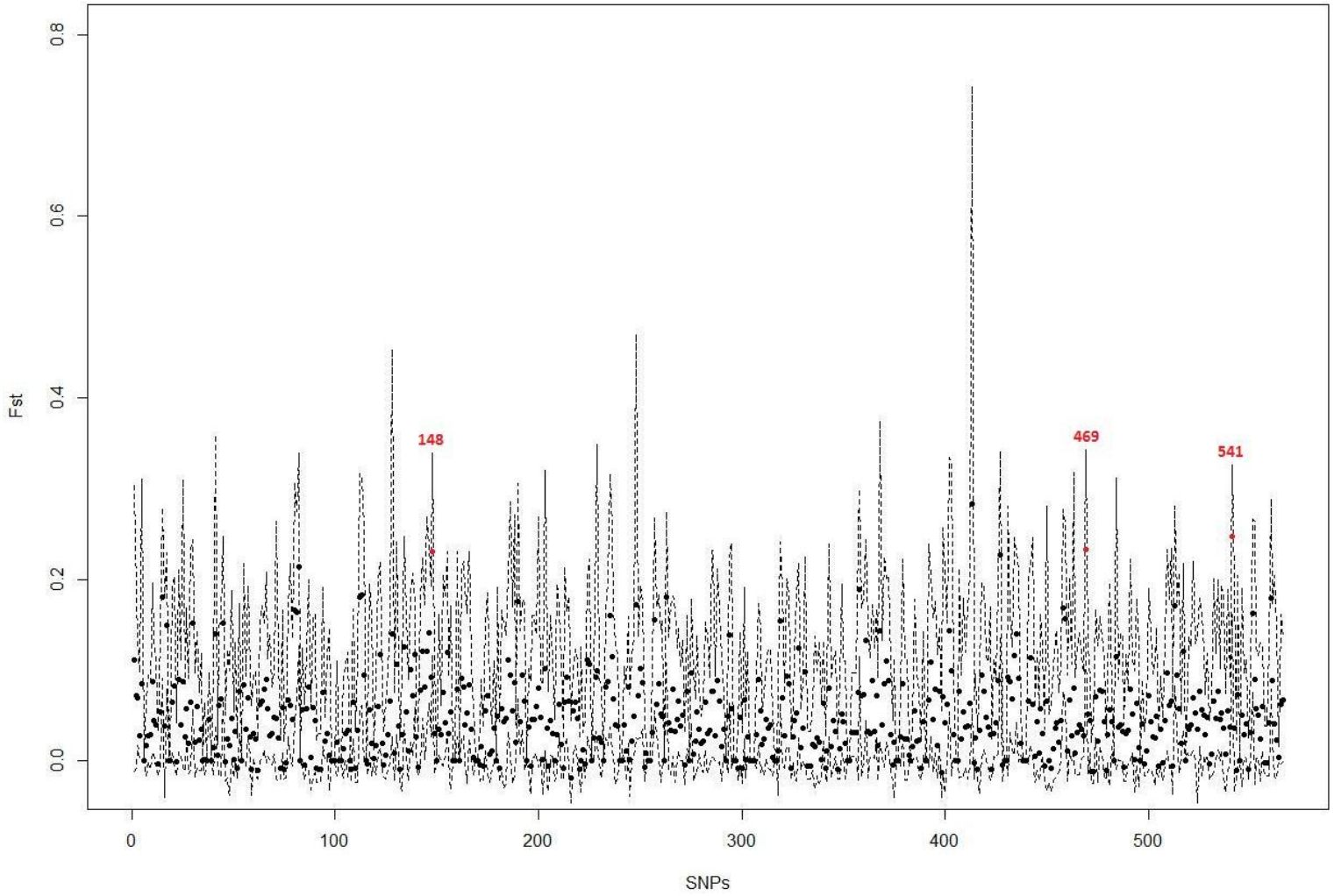


Figure 4.7. Estimations of locus-specific F_{ST} values and 95% CIs for all 566 SNPs. The three consensus outliers are highlighted in red.

For the neutral dataset, global F_{ST} was 0.0269 ($P < 0.001$) and there were 35 significant pairwise comparisons, out of a possible 78 ($P < 0.05$) (Table 4.6). Of these, 15 remained significant after Bonferroni correction ($P < 0.000641$): ISA and NWI; ISA and SWI; ISB and NWI; ISB and SWI; NQ and ISA; Brit and NWI; Hast and NWI; Hast and SWI; Har and ISA; Shet and ISA; LV and NWI; G and ISA; Gul and ISA; Gul and ISB; and Gul and Hast. No additional pairwise comparisons were significant after sequential Bonferroni correction ($P > 0.000794$). Pairwise F_{ST} values revealed no geographical pattern, as illustrated in the corresponding MDS plot (Figure 4.8).

Table 4.6. Neutral dataset: F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Cancer pagurus* in the NE Atlantic. Here, the indicative adjusted nominal level (5%) for multiple comparisons has been calculated as 0.000641 (Bonferroni correction). Significant values at the $P < 0.05$ level are marked in bold. Values remaining significant after Bonferroni correction are marked in bold with an asterisk.

	NWI	**SWI**	**ISA**	**ISB**	**NQ**	**Brit**	**Hast**	**Har**	**Shet**	**LL**	**LV**	**G**	**Gul**
NWI		0.2324	**0.0000***	**0.0000***	0.0723	**0.0000***	**0.0000***	**0.0029**	**0.0107**	**0.0049**	**0.0000***	**0.0440**	0.0801
SWI	0.0264		**0.0000***	**0.0000***	0.4453	0.2549	**0.0000***	0.3369	0.4834	0.1445	0.2022	0.3164	0.3086
ISA	**0.1372***	**0.0904***		0.3184	**0.0000***	**0.0039**	**0.0049**	**0.0000***	**0.0000***	**0.0068**	0.0977	**0.0000***	**0.0000***
ISB	**0.1076***	**0.0630***	0.0187		**0.0166**	**0.0244**	0.0654	**0.0186**	**0.0029**	0.0713	0.0664	**0.0010**	**0.0000***
NQ	0.0385	0.0215	**0.0719***	**0.0442**		0.0645	**0.0010**	0.6436	0.3623	0.5791	0.2090	0.1055	0.5898
Brit	**0.0607***	0.0245	**0.0498**	**0.0382**	0.0357		0.0908	0.2734	0.1846	0.2295	**0.0313**	0.0732	**0.0029**
Hast	**0.1055***	**0.0623***	**0.0455**	0.0334	**0.0536**	0.0304		0.1250	**0.0107**	**0.0117**	0.2744	**0.0039**	**0.0000***
Har	**0.0423**	0.0174	**0.0439***	**0.0305**	0.0116	0.0178	0.0226		0.7783	0.5049	0.3311	0.4356	0.4717
Shet	**0.0343**	0.0140	**0.0540***	**0.0372**	0.0164	0.0192	**0.0343**	0.0057		0.5449	0.3447	0.3828	0.2246
LL	**0.0567**	0.0361	**0.0610**	0.0378	0.0226	0.0302	**0.0528**	0.0187	0.0157		0.5703	0.1377	**0.0068**
LV	**0.1060***	0.0323	0.0376	0.0368	0.0358	**0.0430**	0.0295	0.0226	0.0213	0.0228		0.1611	**0.0039**
G	**0.0464**	0.0267	**0.0921***	**0.0711**	0.0397	0.0374	**0.0633**	0.0196	0.0189	0.0383	0.0404		0.0674
Gul	0.0288	0.0200	**0.0752***	**0.0496***	0.0151	**0.0369**	**0.0563***	0.0109	0.0141	**0.0512**	**0.0583**	0.0366	

For the adaptive dataset, global F_{ST} was 0.1820 ($P < 0.001$) and there were 35 significant pairwise comparisons, out of a possible 78 ($P < 0.05$) (Table 4.7). Of these, nine remained significant after Bonferroni correction ($P < 0.000641$): ISA and NWI; NQ and NWI; NQ and ISB; Shet and NWI; LV and NWI; LV and Har; G and NWI; Gul and NQ; and Gul and LV. No additional pairwise comparisons were significant after sequential Bonferroni correction ($P > 0.000725$).

Table 4.7. Adaptive dataset: F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Cancer pagurus* in the NE Atlantic. Here, the indicative adjusted nominal level (5%) for multiple comparisons has been calculated as 0.000641 (Bonferroni correction). Significant values at the $P < 0.05$ level are marked in bold. Values remaining significant after Bonferroni correction are marked in bold with an asterisk.

	NWI	SWI	ISA	ISB	NQ	Brit	Hast	Har	Shet	LL	LV	G	Gul
NWI		0.0801	**0.0000***	**0.0010**	**0.0000***	**0.0010**	**0.0039**	**0.0068**	**0.0000***	**0.0020**	**0.0000***	**0.0000***	0.3369
SWI	0.1311		**0.0195**	**0.0010**	**0.0215**	**0.0244**	0.0684	0.0645	0.0762	0.3027	**0.0117**	0.1260	0.2207
ISA	**0.4711***	**0.2210**		0.1553	0.0654	0.9990	0.3672	0.1074	0.6143	0.1826	0.0986	0.2305	**0.0010**
ISB	**0.5078**	**0.3788**	0.2258		**0.0000***	0.1543	0.2559	0.0898	**0.0244**	**0.0049**	**0.0010**	**0.0127**	**0.0029**
NQ	**0.5608***	**0.2130**	0.1986	**0.5784***		**0.0440**	**0.0088**	**0.0010**	0.0557	0.6895	0.7813	0.6768	**0.0000***
Brit	**0.4128**	**0.1854**	-0.0365	0.1636	**0.2149**		0.6709	0.2158	0.7295	0.1377	0.0527	0.3027	**0.0098**
Hast	**0.2979**	0.1526	0.0342	0.0652	**0.3106**	-0.0038		0.7627	0.4385	0.0703	**0.0156**	0.1289	0.0723
Har	**0.1881**	0.1080	0.0937	0.0960	**0.3129**	0.0500	-0.0151		0.1592	**0.0322**	**0.0000***	**0.0479**	0.1709
Shet	**0.3233***	0.1046	0.0009	**0.1872**	0.1300	-0.0102	0.0184	0.0451		0.3213	0.0606	0.5420	**0.0088**
LL	**0.4076**	0.0625	0.1366	**0.4819**	-0.0136	0.1333	0.1922	**0.1918**	0.0496		0.4619	0.8555	**0.0098**
LV	**0.6558***	**0.3072**	0.2354	**0.6982**	-0.0186	0.2618	**0.3780**	**0.3695***	0.1711	0.0650		0.4121	**0.0000***
G	**0.4462***	0.1160	0.0448	**0.4281**	-0.0085	0.0540	0.1376	**0.1594**	0.0057	-0.0395	0.0324		**0.0098**
Gul	0.0279	0.0499	**0.2632**	**0.2724**	**0.3974***	**0.2094**	0.1106	0.0433	**0.1545**	**0.2484**	**0.4749***	**0.2639**	

For SNP 148, global F_{ST} was 0.1775 ($P < 0.001$) and there were 22 significant pairwise comparisons, out of a possible 78 ($P < 0.05$) (Table 4.8). Of these, two remained significant after Bonferroni correction ($P < 0.000641$): ISA and NWI; and Brit and NWI. No additional pairwise comparisons were significant after sequential Bonferroni correction ($P > 0.000893$). Pairwise F_{ST} values revealed no geographical pattern, as illustrated in the corresponding MDS plot (Figure 4.9).

Table 4.8. SNP 148: F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Cancer pagurus* in the NE Atlantic. Here, the indicative adjusted nominal level (5%) for multiple comparisons has been calculated as 0.000641 (Bonferroni correction). Significant values at the $P < 0.05$ level are marked in bold. Values remaining significant after Bonferroni correction are marked in bold with an asterisk.

	NWI	**SWI**	**ISA**	**ISB**	**NQ**	**Brit**	**Hast**	**Har**	**Shet**	**LL**	**LV**	**G**	**Gul**
NWI		0.5107	**0.0000***	**0.0010**	**0.0137**	**0.0000***	**0.0107**	**0.0088**	**0.0029**	0.0859	**0.0039**	**0.0215**	0.3252
SWI	-0.0067		**0.0098**	**0.0029**	0.0674	**0.0156**	0.0635	0.0869	**0.0186**	0.3496	**0.0195**	0.0664	0.8594
ISA	**0.6305***	**0.4783**		0.9990	0.4951	0.9990	0.4580	**0.0342**	0.2900	0.1533	0.9990	0.1367	**0.0029**
ISB	**0.6305**	**0.4783**	0.0000		0.4463	0.9990	0.4629	**0.0322**	0.2754	0.1563	0.9990	0.1455	**0.0039**
NQ	**0.4125**	0.2500	0.0870	0.0870		0.6973	0.9990	0.4629	0.9990	0.5371	0.4678	0.9990	**0.0498**
Brit	**0.5541***	**0.3949**	0.0000	0.0000	0.0029		0.7363	0.1494	0.4014	0.3447	0.9990	0.5117	**0.0117**
Hast	**0.4125**	0.2500	0.0870	0.0870	-0.0435	0.0029		0.4424	0.9990	0.5303	0.4854	0.9990	0.0537
Har	**0.2766**	0.1274	**0.1565**	**0.1565**	0.0029	0.0902	0.0029		0.5430	0.9990	0.1338	0.6035	0.1084
Shet	**0.4138**	**0.2511**	0.0853	0.0853	-0.0304	0.0230	-0.0304	0.0017		0.5830	0.4356	0.9990	**0.0293**
LL	0.2176	0.0740	0.2558	0.2558	0.0007	0.1351	0.0007	-0.0422	-0.0042		0.4551	0.7129	0.3711
LV	**0.5795**	**0.4252**	0.0000	0.0000	0.0575	-0.0180	0.0575	0.1307	0.0642	0.2000		0.4824	**0.0225**
G	**0.3880**	0.2264	0.1052	0.1052	-0.0550	-0.0034	-0.0550	-0.0089	-0.0416	-0.0133	0.0667		0.0830
Gul	0.0344	-0.0294	**0.3770**	**0.3770**	**0.1839**	**0.3070**	0.1839	0.0811	**0.1836**	0.0360	**0.3375**	0.1653	

For SNP 469, global F_{ST} was 0.1682 ($P < 0.001$) and there were 24 significant pairwise comparisons, out of a possible 78 ($P < 0.05$) (Table 4.9). Of these, seven remained significant after Bonferroni correction ($P < 0.000641$): NQ and NWI; NQ and ISB; Har and NQ; LV and NWI; LV and Har; Gul and NQ; and Gul and LV. No additional pairwise comparisons were significant after sequential Bonferroni correction ($P > 0.000704$). Pairwise F_{ST} values revealed no geographical pattern, as illustrated in the corresponding MDS plot (Figure 4.10).

Table 4.9. SNP 469: F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Cancer pagurus* in the NE Atlantic. Here, the indicative adjusted nominal level (5%) for multiple comparisons has been calculated as 0.000641 (Bonferroni correction). Significant values at the $P < 0.05$ level are marked in bold. Values remaining significant after Bonferroni correction are marked in bold with an asterisk.

	NWI	**SWI**	**ISA**	**ISB**	**NQ**	**Brit**	**Hast**	**Har**	**Shet**	**LL**	**LV**	**G**	**Gul**
NWI		**0.0371**	0.0879	0.9990	**0.0000***	0.1055	0.3604	0.3057	**0.0176**	**0.0010**	**0.0000***	**0.0098**	0.4209
SWI	**0.2719**		0.9990	**0.0469**	0.1133	0.8418	0.3076	0.1357	0.9990	0.3398	0.1172	0.4736	0.0879
ISA	0.2258	-0.0398		0.1621	0.0869	0.9990	0.4941	0.2969	0.9990	0.2490	0.0898	0.6504	0.1934
ISB	-0.0435	**0.2719**	0.2258		**0.0000***	0.1816	0.3613	0.3555	**0.0195**	**0.0078**	**0.0010**	**0.0215**	0.4648
NQ	**0.6611***	0.1777	0.2240	**0.6611***		0.0615	**0.0068**	**0.0000***	**0.0420**	0.7656	0.9990	0.4951	**0.0000***
Brit	0.1808	-0.0287	-0.0397	0.1808	0.2727		0.6074	0.3975	0.8438	0.2227	0.0537	0.3906	0.2754
Hast	0.0559	0.0510	0.0198	0.0559	**0.4306**	-0.0062		0.9990	0.3115	**0.0430**	**0.0107**	0.1055	0.8115
Har	0.0393	0.0877	0.0519	0.0393	**0.4621***	0.0211	-0.0310		0.1270	**0.0176**	**0.0000***	**0.0400**	0.8281
Shet	**0.2251**	-0.0314	-0.0313	**0.2251**	**0.1982**	-0.0240	0.0423	0.0738		0.2803	0.0762	0.4131	0.0625
LL	**0.5583**	0.0513	0.0895	**0.5583**	-0.0248	0.1324	**0.2888**	**0.3355**	0.0787		0.7822	0.9990	**0.0059**
LV	**0.6982***	0.1882	0.2354	**0.6982**	-0.0535	0.2852	**0.4489**	**0.4804***	0.2090	-0.0231		0.5527	**0.0000***
G	**0.4862**	0.0023	0.0329	**0.4862**	0.0161	0.0694	0.2137	**0.2623**	0.0276	-0.0575	0.0222		**0.0195**
Gul	0.0113	0.1319	0.0911	0.0113	**0.5182***	0.0542	-0.0229	-0.0182	0.1107	**0.3956**	**0.5396***	**0.3215**	

For SNP 541, global F_{ST} was 0.1989 (P < 0.001) and there were 24 significant pairwise comparisons, out of a possible 78 (P < 0.05) (Table 4.10). Of these, seven remained significant after Bonferroni correction (P < 0.000641): ISA and NWI; ISA and SWI; NQ and NWI; Har and ISA; Shet and ISA; G and ISA; and Gul and ISA. No additional pairwise comparisons were significant after sequential Bonferroni correction (P > 0.000704). Pairwise F_{ST} values revealed no geographical pattern, as illustrated in the corresponding MDS plot (Figure 4.11).

Table 4.10. SNP 541: F_{ST} pairwise comparisons (lower diagonal) and the associated P-values (upper diagonal) for *Cancer pagurus* in the NE Atlantic. Here, the indicative adjusted nominal level (5%) for multiple comparisons has been calculated as 0.000641 (Bonferroni correction). Significant values at the P < 0.05 level are marked in bold. Values remaining significant after Bonferroni correction are marked in bold with an asterisk.

	NWI	**SWI**	**ISA**	**ISB**	**NQ**	**Brit**	**Hast**	**Har**	**Shet**	**LL**	**LV**	**G**	**Gul**
NWI		0.7656	**0.0000***	**0.0010**	**0.0000***	0.0684	0.1856	0.0781	0.1377	0.1221	**0.0195**	0.9990	0.5098
SWI	-0.0159		**0.0000***	**0.0059**	**0.0156**	0.2510	0.5010	0.2617	0.3242	0.3438	0.0693	0.5381	0.9990
ISA	**0.8387***	**0.7214***		0.3594	0.1494	**0.0039**	**0.0020**	**0.0000***	**0.0000***	**0.0049**	**0.0449**	**0.0000***	**0.0000***
ISB	**0.6103**	**0.4864**	0.0811		0.7979	0.0811	**0.0479**	**0.0147**	**0.0156**	0.1455	0.4190	**0.0029**	**0.0020**
NQ	**0.5026***	**0.3762**	0.1676	-0.0243		0.2647	0.1602	0.0654	0.0586	0.2295	0.7822	**0.0029**	**0.0010**
Brit	0.2083	0.0936	**0.4412**	0.1777	0.0826		0.8125	0.6397	0.6738	0.9990	0.4600	0.0811	0.1025
Hast	0.1224	0.0244	**0.5343**	**0.2727**	0.1665	-0.0287		0.9990	0.9990	0.8223	0.3320	0.1504	0.3926
Har	0.1323	0.0427	**0.4697***	**0.2502**	0.1547	-0.0232	-0.0322		0.9990	0.8174	0.2041	0.0703	0.1445
Shet	0.1173	0.0316	**0.4857***	**0.2692**	0.1727	-0.0174	-0.0323	-0.0212		0.8311	0.2549	0.1250	0.2520
LL	0.1930	0.0692	**0.5002**	0.1998	0.0947	-0.0540	-0.0464	-0.0393	-0.0349		0.5664	0.1250	0.2139
LV	**0.4136**	0.2732	**0.2838**	0.0161	-0.0384	0.0023	0.0694	0.0648	0.0805	0.0061		**0.0176**	**0.0273**
G	-0.0579	0.0123	**0.8816***	**0.6496**	**0.5421**	0.2508	0.1643	0.1688	0.1532	0.2416	**0.4595**		0.5147
Gul	-0.0120	-0.0368	**0.7056***	**0.4934**	**0.3890**	0.1067	0.0348	0.0527	0.0412	0.0823	**0.2892**	0.0135	

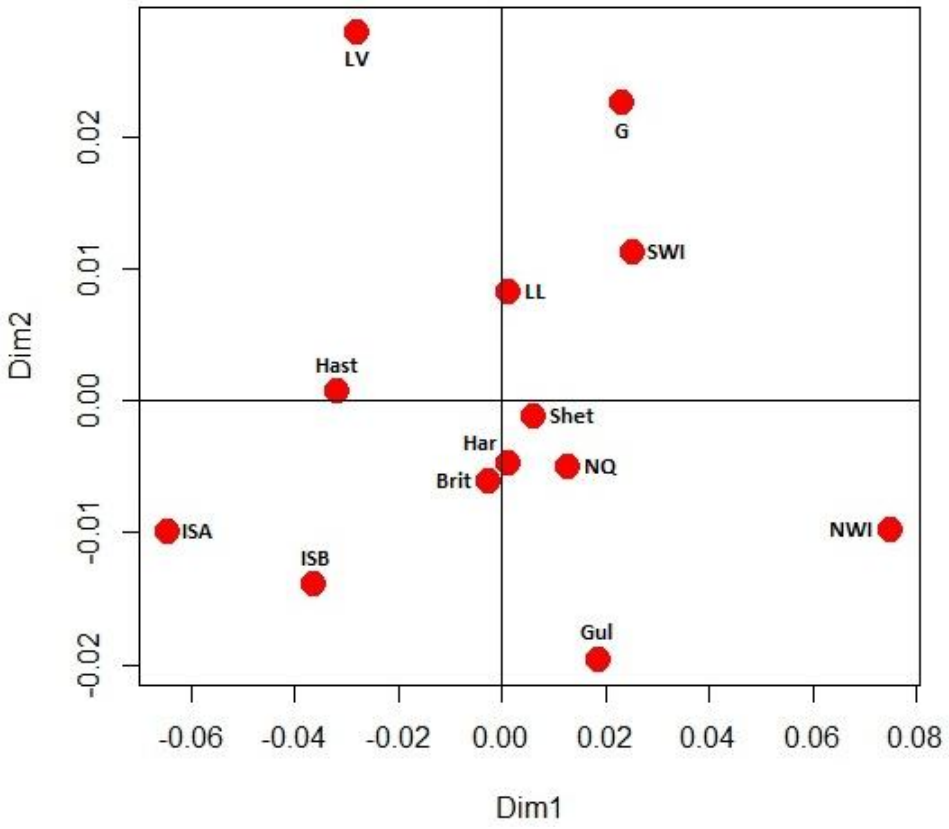


Figure 4.8. Neutral dataset: MDS plot of pairwise F_{st} values of *Cancer pagurus*.

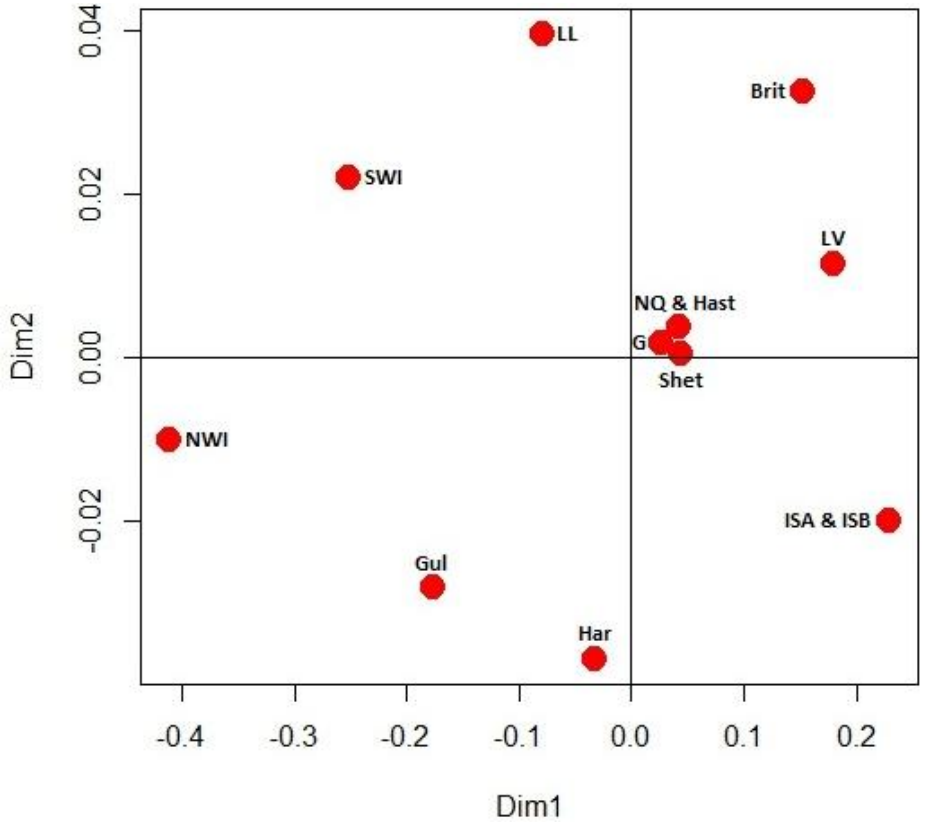


Figure 4.9. SNP 148: MDS plot of pairwise F_{st} values of *Cancer pagurus*.

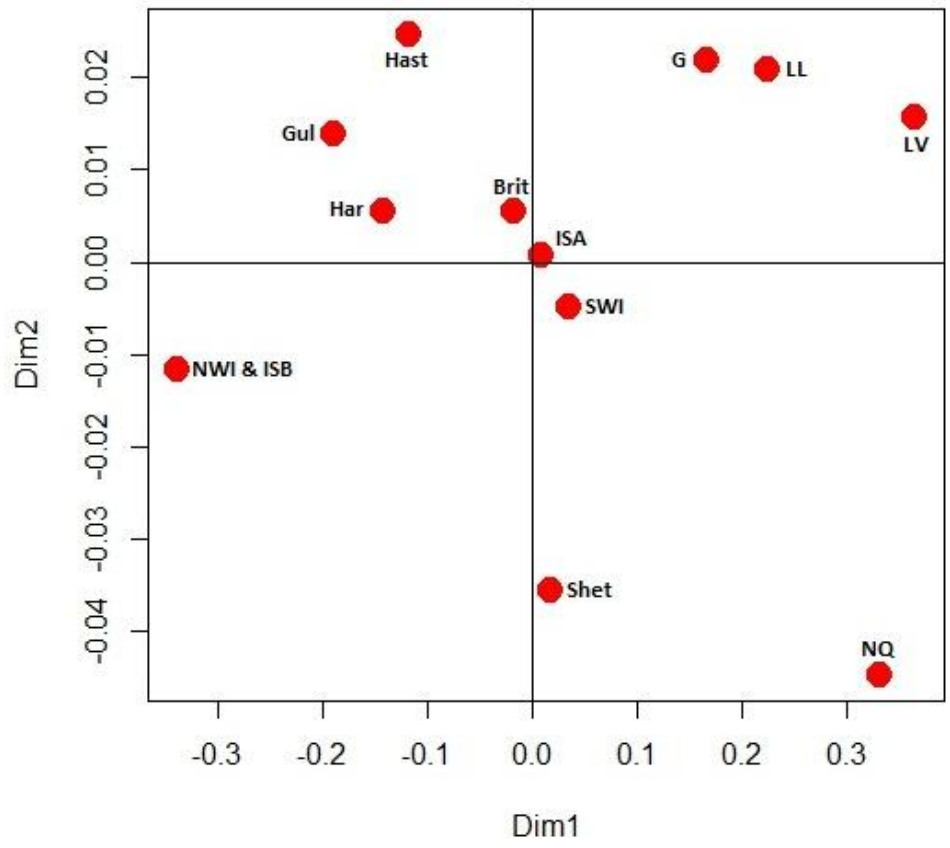


Figure 4.10. SNP 469: MDS plot of pairwise F_{st} values of *Cancer pagurus*.

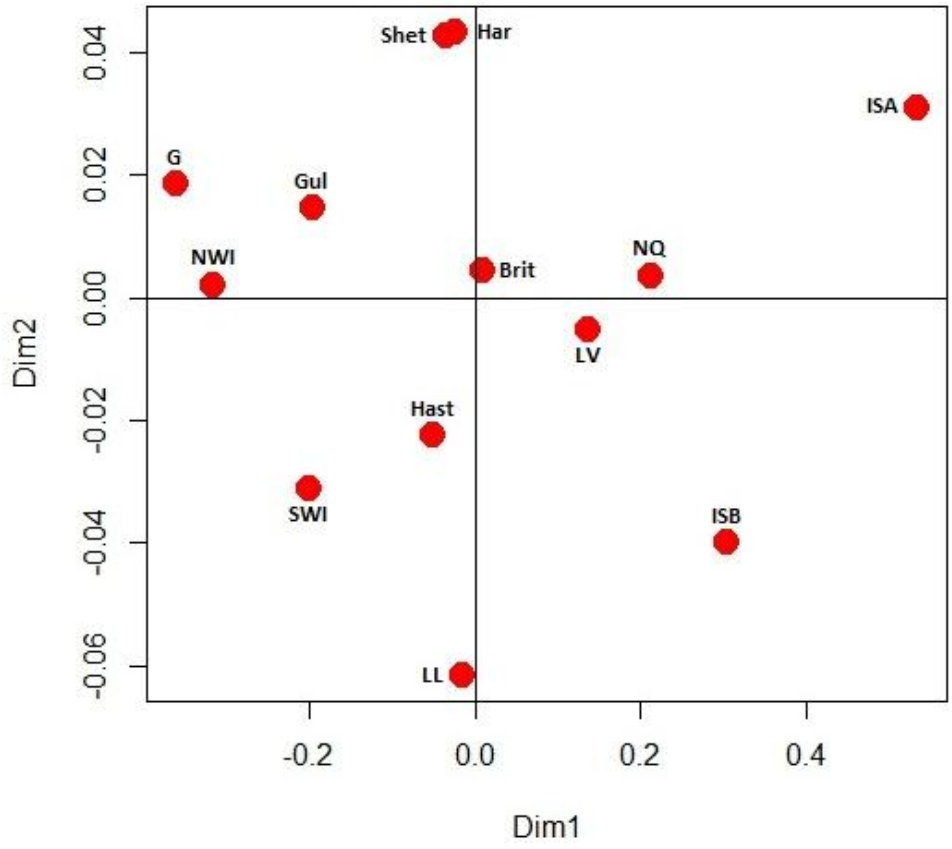


Figure 4.11. SNP 541: MDS plot of pairwise F_{st} values of *Cancer pagurus*.

The mantel test for matrix correlation between genetic and geographic distances revealed no evidence of IBD in the neutral dataset (r = -0.0450, P = 0.6424) (Figure 4.12). Similarly, IBD was not detected for the three positive outlier loci: SNP 148 (r = -0.1058, P = 0.8588) (Figure 4.13); SNP 469 (r = -0.0192, P = 0.4904) (Figure 4.14); and SNP 541 (r = -0.0011, P = 0.4343) (Figure 4.15).

PCA for both the neutral and adaptive datasets failed to reveal any geographically coherent patterns (neutral: Figure 4.16, adaptive: Figure 4.17).

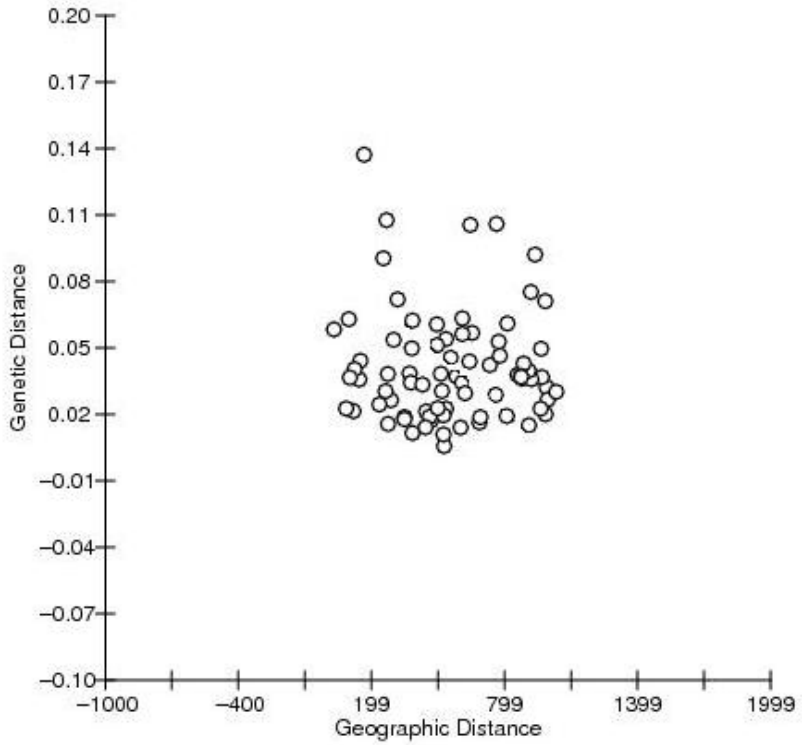


Figure 4.12. Neutral dataset: Testing for IBD in *Cancer pagurus* in the NE Atlantic, with genetic distance (F_{ST}) and geographic distance (km).

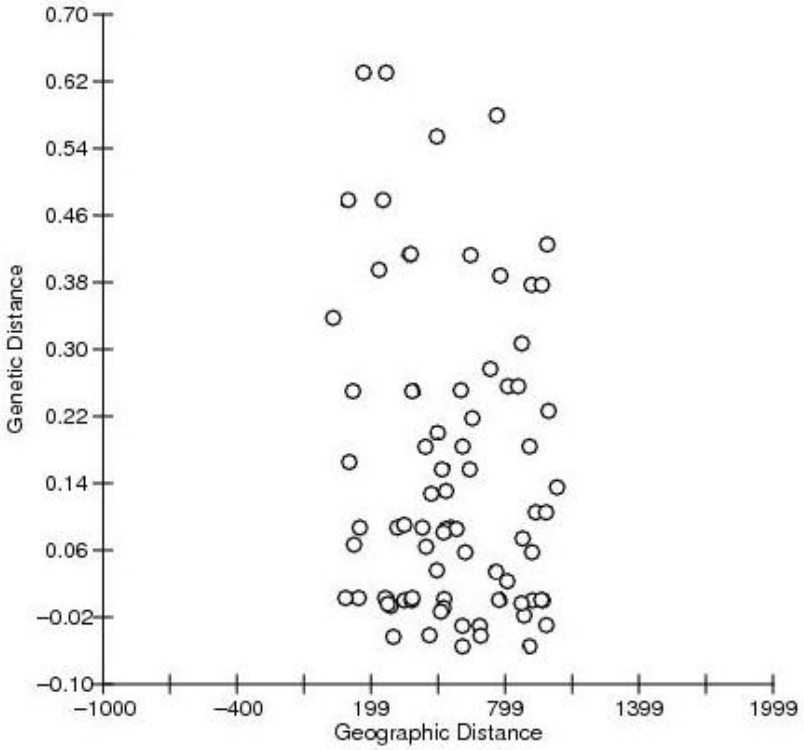


Figure 4.13. SNP 148: Testing for IBD in *Cancer pagurus* in the NE Atlantic, with genetic distance (F_{ST}) and geographic distance (km).

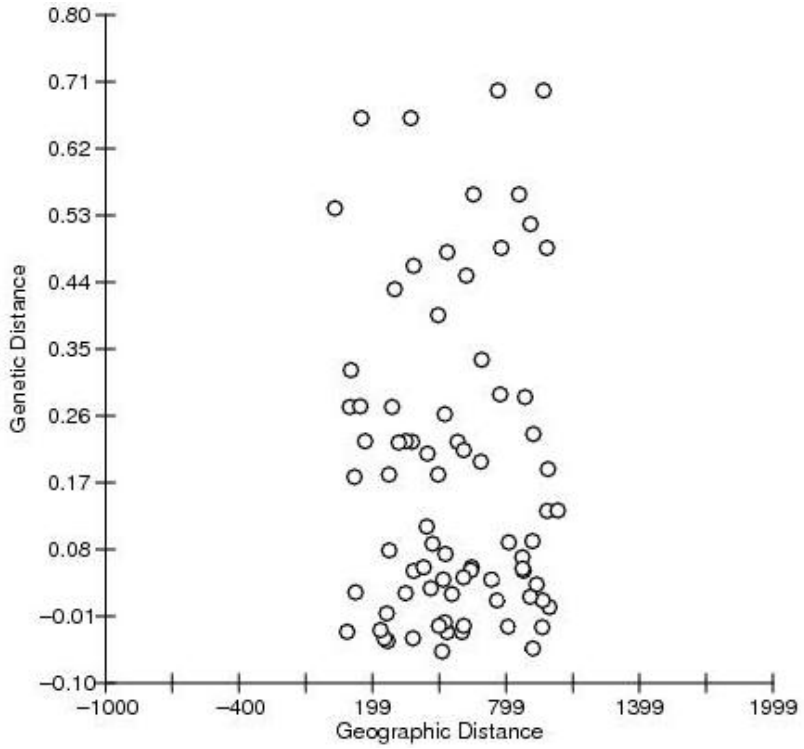


Figure 4.14. SNP 469: Testing for IBD in *Cancer pagurus* in the NE Atlantic, with genetic distance (F_{ST}) and geographic distance (km).

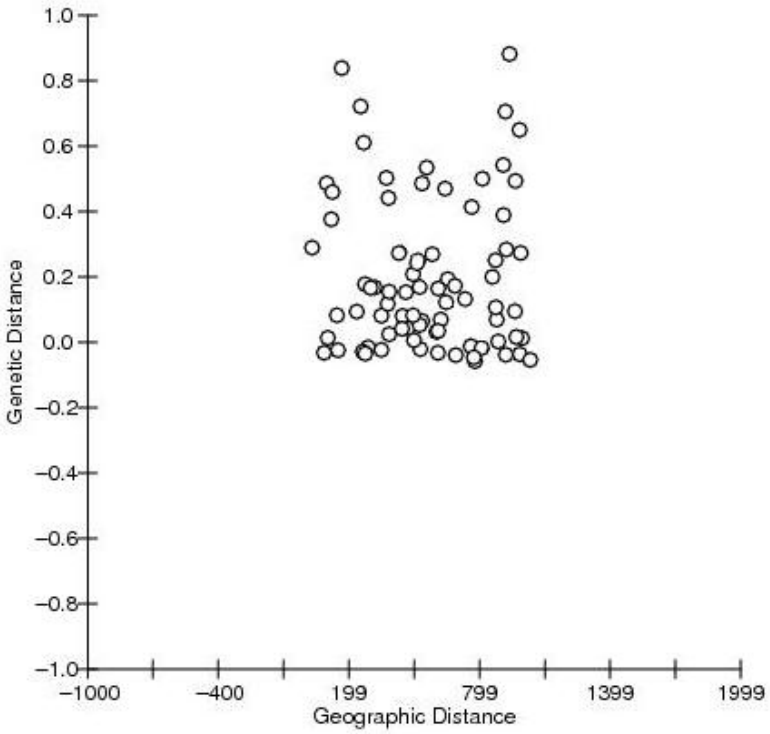


Figure 4.15. SNP 541: Testing for IBD in *Cancer pagurus* in the NE Atlantic, with genetic distance (F_{ST}) and geographic distance (km).

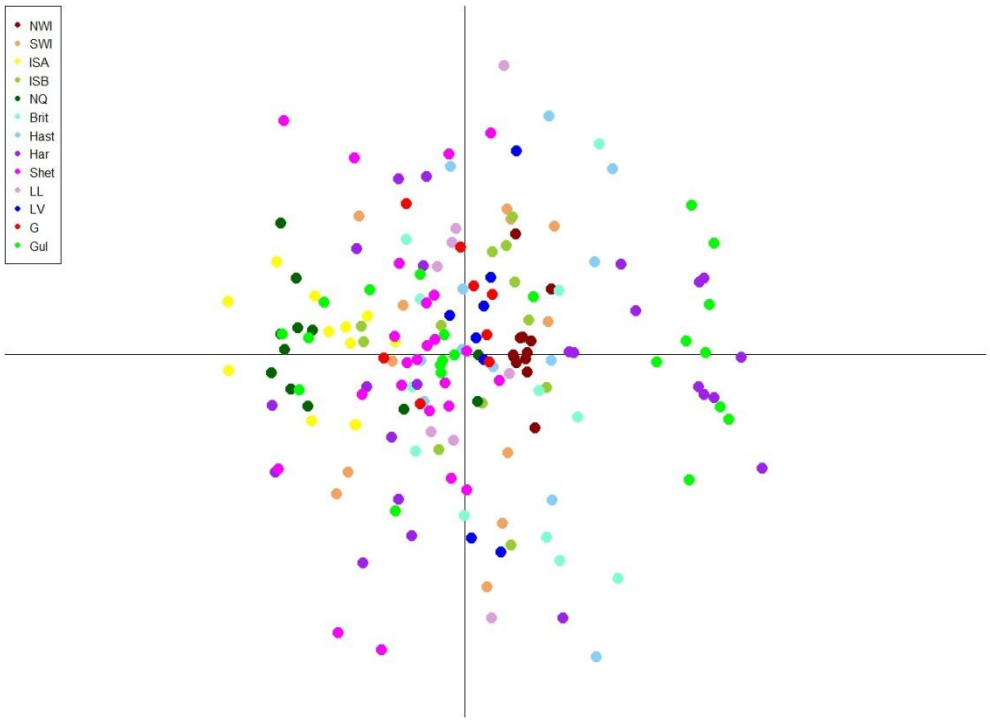


Figure 4.16. PCA for the neutral dataset.

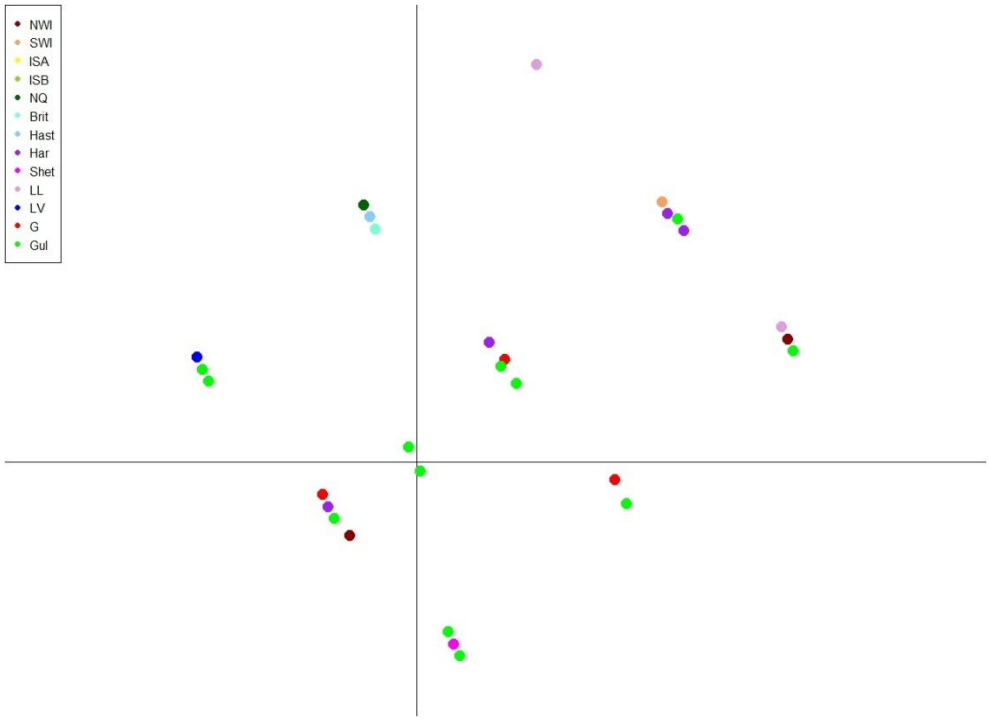


Figure 4.17. PCA for the adaptive dataset.

A mean likelihood of K = 1 (LnP(K): loc prior, -39847.5; no loc prior, -39847.5) (Figure 4.18, loc prior; Figure 4.19, no loc prior) gives no support for population structuring in the neutral dataset (Figure 4.20, loc prior; Figure 4.21, no loc prior).

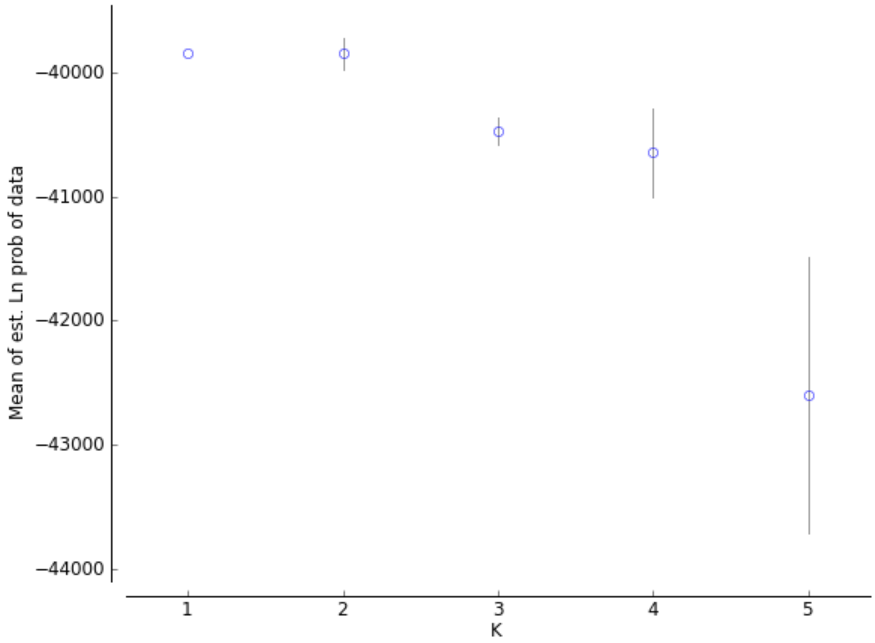


Figure 4.18. Plot of mean likelihood L(K) and variance per K value from STRUCTURE; loc prior.

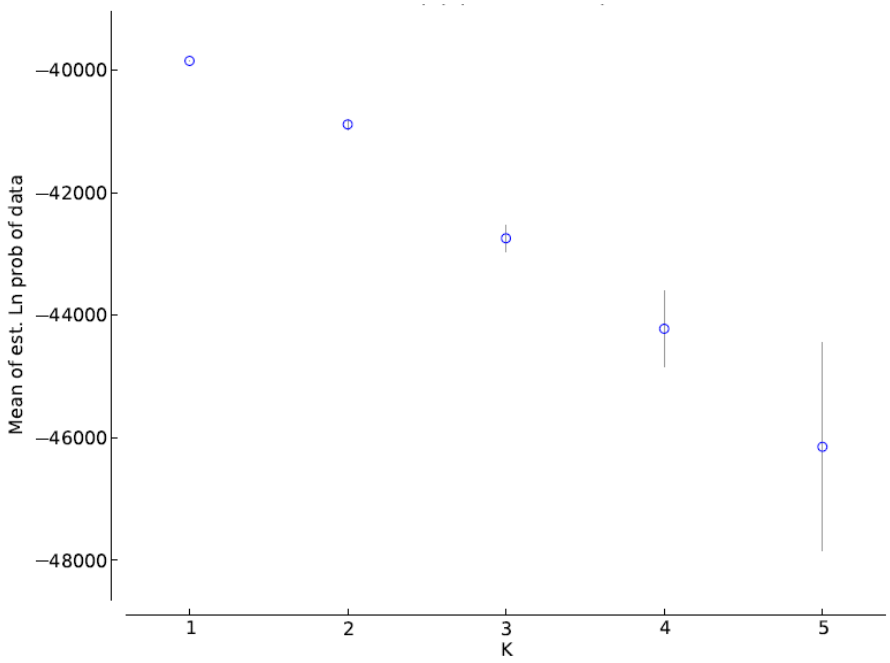


Figure 4.19. Plot of mean likelihood L(K) and variance per K value from STRUCTURE; no loc prior.

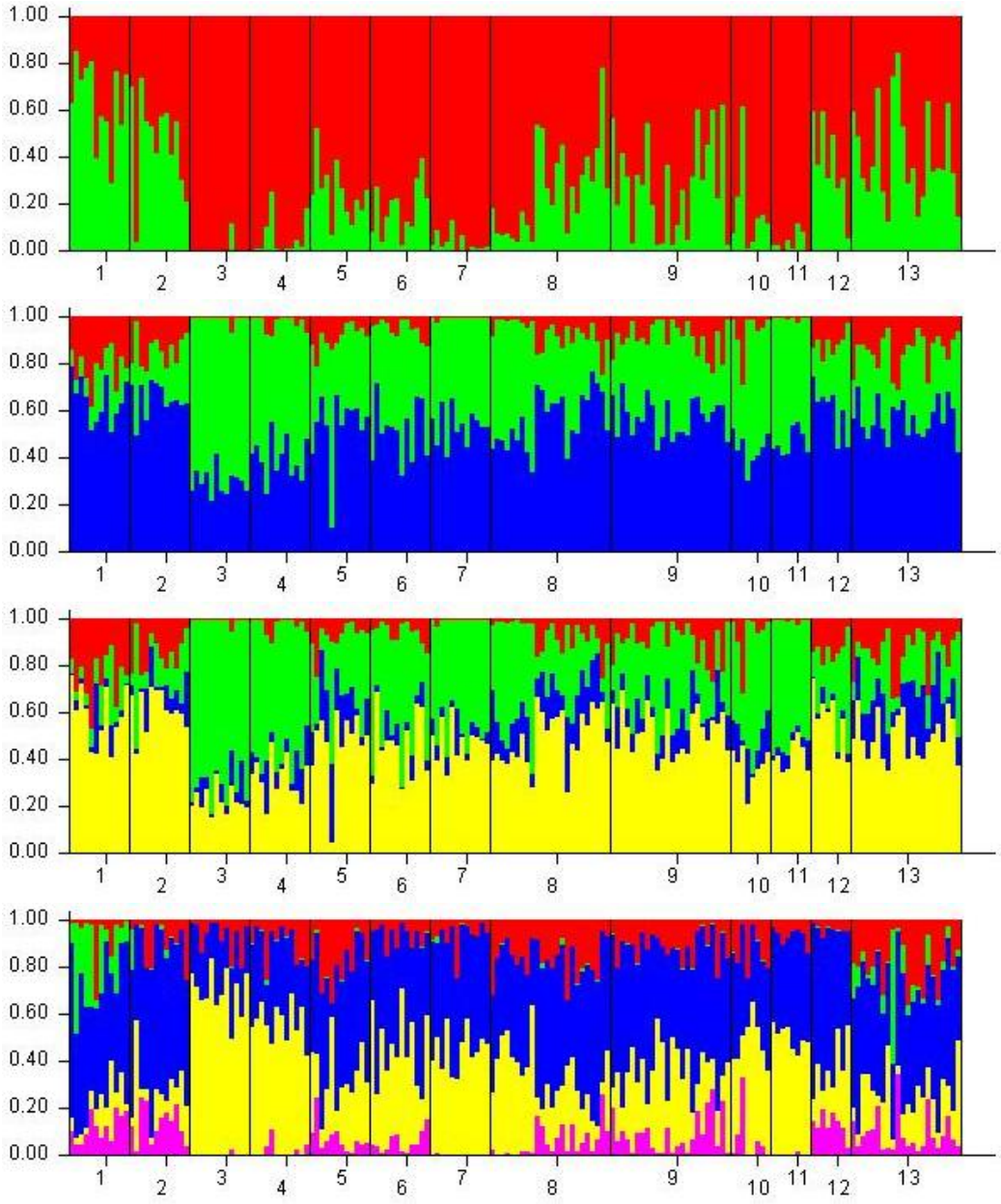


Figure 4.20. STRUCTURE bar plots for *Cancer pagurus* for K = 2 to 5; loc prior analysis. Each vertical bar represents an individual and each colour a different cluster.

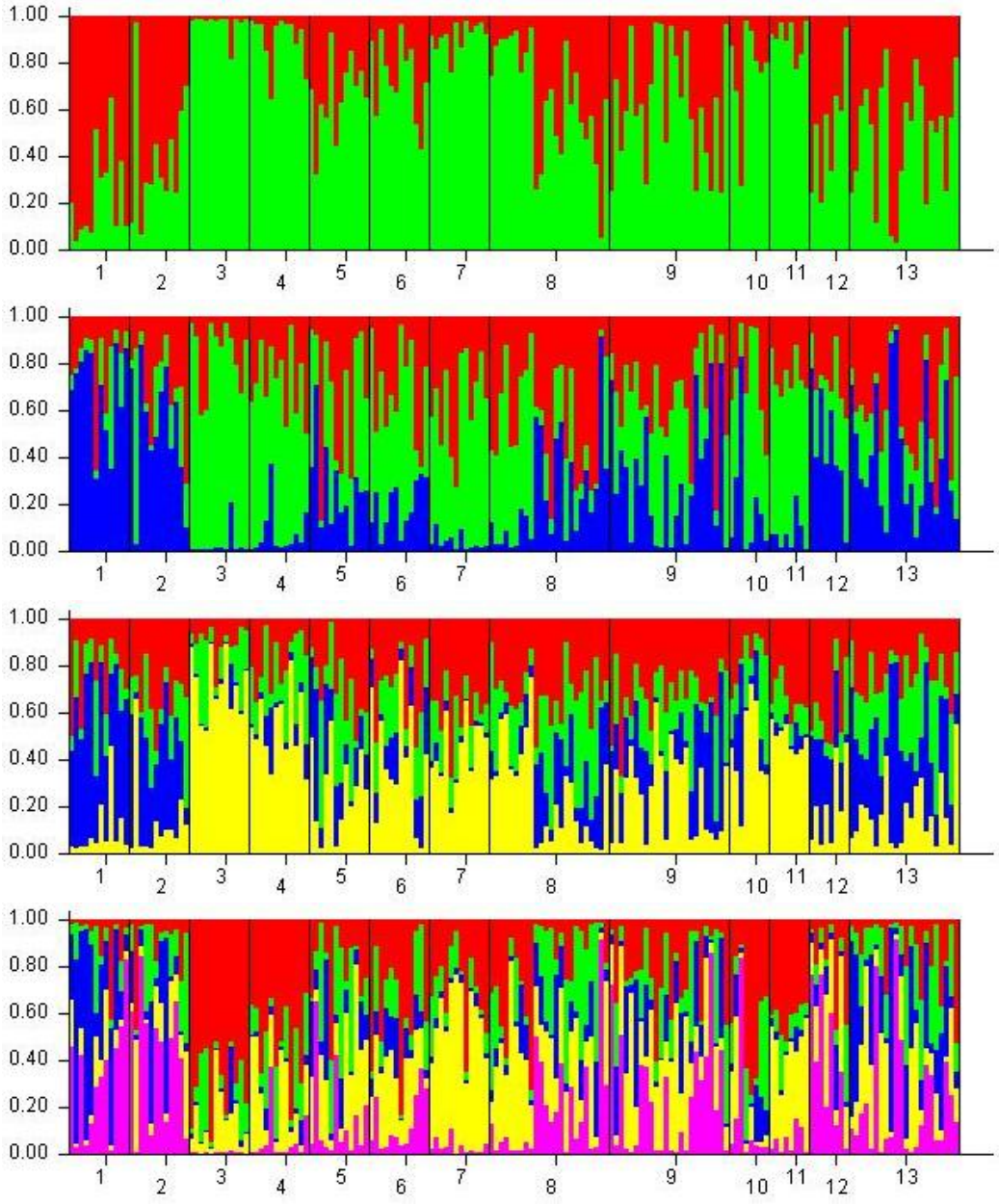


Figure 4.21. STRUCTURE bar plots for *Cancer pagurus* for K = 2 to 5; no loc prior analysis. Each vertical bar represents an individual and each colour a different cluster.

4.4. DISCUSSION

The present study combined exploratory RAD sequencing together with a spatial sampling scheme that encompassed a large portion of the distributional range of *C. pagurus*. The fundamental objective was to use this approach to assay a large number of loci to provide a deeper understanding of the factors shaping the genetic population structure of the species, and an insight into the utility of this approach for future studies of this species and region.

The number of SNPs recovered (N = 566) was much lower than numbers identified in studies of other taxa focusing on smaller geographic areas. For example, 5985 novel SNPs were detected in two Baltic Sea populations of herring, *Clupea harengus*, separated by 387 km (Corander et al. 2013). While imposing less stringent SNP filtering criteria considerably increased the number of SNPs detected it also dramatically increased the number of missing genotypes. Poor quality DNA is likely to have been a contributing factor, as most DNA was extracted through phenol chloroform and had been stored at -20 °C for a considerable period of time (<10 years). Another issue serving to reduce the number of SNPs was the absence of a reference genome. 132 SNPs with three and four alleles were identified. However, as most SNPs are expected to be biallelic, these were subsequently discounted. While low levels of drift in species like *C. pagurus* may serve to maintain rare alleles, it is likely that these SNPs were due to paralogous sequences, although alignment with a scaffold reference genome is needed to confirm this. The analysed suite of SNPs thus represents a trade-off between SNP number and both genotype robustness (as measured by STACK depth: minimum per individual allele = 30) and sample coverage (less than 5% missing data per individual). The SNPs were initially analysed using multiple marker-based neutrality tests to partition those SNPs conforming to neutral expectation from those identified as positive outliers that exceeded neutral expectations of genetic structuring. The results for both 'neutral' and 'adaptive' SNPs are discussed separately.

4.4.1. 'Neutral' Dataset

When comparing the resolving power of nuclear SNPs with other types of markers, such as microsatellites, it is expected that the use of 4-12 SNPs confers similar

resolving power to a single microsatellite locus (Guichoux et al. 2011), assuming neutrality. As such, the neutral SNP dataset offers enhanced power to detect population divergences resulting from genetic drift compared to the previous microsatellite study. In line with this expectation, and the fact that F_{ST} values derived from microsatellites are typically deflated compared to other loci due to the high levels of heterozygosity (Hedrick 2005), the neutral SNP data set yielded F_{ST} values that were an order of magnitude higher than values derived from microsatellites (Chapter 3) (global RAD F_{ST} = 0.0278, $P < 0.001$; global microsatellite F_{ST} = 0.003, 95% CI: 0.002 – 0.005). In addition, although STRUCTURE supported a model of K = 1, there were a large number of significant pairwise tests. This discordance may reflect the fact that STRUCTURE has been shown to lack resolution at low levels of inter-population divergence (Latch et al. 2006) and, more specifically, that a maximum allowable deviation from HWE was imposed as part of the SNP calling procedure, which may limit the downstream power of this analysis. Genetic structuring did not conform to an IBD model or an obvious hierarchical geographic pattern, with geographically close samples often more differentiated than more distant samples. The pattern revealed by the neutral SNPs therefore represents a more extreme case of chaotic genetic patchiness to that reported in Chapter 3.

There is a consensus in most studies showing genetic patchiness on a microgeographical scale that such heterogeneity likely stems from spatial/temporal genetic variation in the composition of recruits generated by large variances in reproductive success (Johnson & Black 1982; Hedgecock 1994; Li & Hedgecock 1998; Planes & Lenfant 2002). As described in Chapter 3, female sperm usage patterns and harvesting practises may amplify such variance effects in the case of *C. pagurus*. Non-mutually exclusive alternative hypotheses include natural selection (Nielsen et al. 2009), population isolation within a non-migration-drift equilibrium system (Stamatis et al. 2004), and variable recruitment from genetically distinct populations (Selkoe et al. 2006). A fuller insight into the mechanisms driving the chaotic genetic patchiness will require further analysis of genetic variation

throughout the geographic range (spatial variation), and over multiple years or different stages (e.g. adults, juveniles or larvae; temporal variation).

4.4.2. 'Adaptive' Dataset

Demonstrating local adaptation in the wild is a notoriously difficult task (Gienapp et al. 2008), however, as a step towards this there is a large interest in identifying the 'molecular signals of local adaptation' (Vasemägi & Primmer 2005). Such enterprises are typically grouped under the term 'genomics' which broadly defined, refers to the population genetic analyses of a large number of loci with view to separating genome wide (neutral) and locus specific (selection) effects. For taxa like *C. pagurus* where genomic resources are limited, RADseq represents a method of genotyping a large number of loci and increasing power to detect molecular signatures of adaptation. In this study, the combination of FDIST and BAYESCAN multiple marker neutrality tests were employed as these two methods have been shown to have the lowest type II (false negative) error, whilst BAYESCAN has the lowest type I (false positive) error (Narum & Hess 2011) among currently employed outlier tests. In line with this, global and pairwise FDIST tests consistently reported a greater number of positive outliers compared to the comparable BAYESCAN tests. Overall global analysis revealed three consensus outliers between both methods, while there were no consensus outliers between methods identified in any pairwise tests which are predicted to yield more reliable results (Robertson 1975).

The number of identified positive outliers is considerably less than those reported in other genome scan-like studies of marine taxa. For example, Corander et al. (2013) identified 4756 polymorphic SNPs between *C. harengus* populations and highlighted 117 as showing evidence for substantial divergence. Previously, Limborg et al. (2012) found statistically strong evidence for divergent selection at fewer loci than Corander et al. (2013) in *C. harengus*, but still more than the present study; 281 SNPs, with 265 neutral and 16 outlier loci. The aforementioned studies also reported clear spatial patterns of variation at outlier loci that paralleled prominent environmental gradients. In contrast, allele frequencies at the 3 outlier SNPs identified here did not show a clear spatial pattern, with the exception of SNP 541 which suggested differentiation of both Irish Sea samples from other samples.

The failure to detect a robust environmentally coherent signal of local adaptation is perhaps not that unexpected given that while there are an increasing number of studies reporting molecular signatures of local adaptation among NE Atlantic species (Hemmer-Hansen et al. 2007; Milano et al. 2014), these have primarily revealed such divergence between the Baltic and Atlantic waters. However, that is not to say that local adaptations are not present in *C. pagurus*. It is possible that the within region heterogeneity may be less pronounced, corresponding with reduced selection pressures and, therefore, less conspicuous patterns. The neutral chaotic genetic patchiness detected for *C. pagurus* is an important consideration for both the potential detection and accumulation of local adaptations. Firstly, by generating increased levels of background genetic differentiation between samples such patchiness will reduce the power of multiple marker based neutrality tests, that are based on construction of a null model, to identify positive outliers (Beaumont 2005). Furthermore, the surprising levels of adaptation in the face of high levels of gene flow reported for marine taxa are believed to be facilitated by low levels of genetic drift (Nielsen et al. 2009). The recurrent, albeit potentially ephemeral, genetic drift driving such patchiness may serve to hinder the development of locally adapted gene complexes in *C. pagurus*.

While it was not possible to identify environmental drivers of variation for the outlier loci, their detection demonstrates the potential of RADseq to provide highly informative SNP loci for stock discrimination/fish forensics. Loci under even mild selection pressures may attain equilibrium much more quickly than neutral loci (Canino et al. 2005). While such levels of selection may be important over evolutionary timescales, they may be insignificant on timescales of interest to fishery managers. Therefore, instead of biasing their application as population markers, loci under selection may actually be more useful in describing population structure in many marine populations. For example, greater structuring at positive outlier SNPs compared to neutral markers has been reported in *C. harengus* (Limborg et al. 2012; Corander et al. 2013). Limborg et al. (2012) compared two datasets generated from 607 individuals from 18 spawning locations in the Northeast Atlantic: one including both outlier and neutrally behaving SNP loci (the

'full' marker set) and one where all outlier loci, as detected in global tests, were removed (the 'neutral' marker set). The 'full' marker set identified four major groups of *C. harengus*, namely the Baltic Sea, Baltic-North Sea transition, North Sea/British Isles, and North Atlantic. In contrast, the 'neutral' marker set recognised three clusters: the Baltic Sea, Baltic/North Sea transition area, and the North Sea/British Isles/North Atlantic. This highlights the increased ability of distinguishing weakly structured populations by including a small number of loci influenced by selection. Likewise, Corander et al. (2013) reported considerably higher levels of differentiation at outlier SNPs among *C. harengus* populations (F_{ST} = 0.128 (95% CI: 0.125 – 0.131)) compared to estimates derived from other marker types, such as F_{ST} = 0.001-0.009 with allozymes (Ryman et al. 1984; André et al. 2011) and F_{ST} = 0.002 with microsatellites (Jørgensen et al. 2005; André et al. 2011), showing that even loci affected by cryptic divergent selection can serve as powerful population markers in high gene flow species.

5. GENERAL DISCUSSION

5.1. Main Findings

- Intensive fishing can have negative ecological consequences on population growth and the structure of food webs (Hutchings & Reynolds 2004), as well as leading to evolutionary changes over short time scales (Olsen et al. 2004) including the loss of genetic diversity (Hauser et al. 2002; Hutchinson et al. 2003). Such genetic erosion may compromise adaptability, population viability, and productivity (Hauser et al. 2002). Accordingly, a fundamental objective of this research was to assess genetic variability and structuring in two commercially harvested decapods, the European lobster (*Homarus gammarus*) and edible crab (*Cancer pagurus*), with a focus on the Irish Sea.
- Within the Irish Sea levels of genetic variability were high for both species. Furthermore, there was no evidence of a recent bottleneck and N_e estimates were infinitely large. Large effective population sizes have also been reported for the European spiny lobster *Palinurus elephas* (Palero et al. 2011) while much lower values have been reported in overfished taxa such as the New Zealand snapper *Pagrus auratus* (Hauser et al. 2002), northern pike *Esox lucius* (Miller & Kapuscinski 1997) and North Sea cod *Gadus morhua* (Hutchinson et al. 2003).
- Population structure analysis indicated no significant structuring for either species among analysed samples in the Irish Sea, supporting the view that *H. gammarus* and *C. pagurus* within the sampled region are derived from single panmictic, at least on evolutionary timescales, populations. This is compatible with studies indicating high gene flow throughout Northeast Atlantic waters revealed in other studies of *H. gammarus* (Triantafyllidis et al. 2005) and *C. pagurus* (Ungfors et al. 2009), and other crustaceans with high dispersal potential (Duran et al. 2004; Couceiro et al. 2007; Sotelo et al. 2008; Sotelo et al. 2009; Domingues et al. 2010).
- For both species, samples collected from the Lundy NTZ exhibited significantly higher positive F_{IS} values. Such values occurring against a background of high

gene flow are best explained by increased variance in reproductive success at the NTZ generating within and/or between cohort genetic differences.

- As biophysical modelling of lobster larvae indicate the majority of recruits are derived from allochthonous larvae, such variance in reproductive success likely stems from post-settlement processes within the NTZ, as opposed to the occurrence of larger more fecund females within the NTZ.
- Extending the genetic survey of *C. pagurus* to include samples from throughout the Northeast Atlantic, both microsatellite and RADseq analysis indicated a lack of large scale geographic structure consistent with extensive spatial gene flow upon which there was evidence of chaotic genetic patchiness. Such patchiness was more pronounced for the RADseq dataset, likely due to the greater sensitivity to genetic structuring conferred by the large number of loci.
- Although natural selection and variability in recruitment patterns from genetically distinct populations may contribute to the genetic patchiness, it is suggested that genetic drift associated with large variances in reproductive success among individuals plays a prominent role. Spatial variability in the extent of variance in reproductive success may be influenced by directional fishing pressures (non-random removal of large females), as well as oceanography. For example low N_e estimates reported for the Galway Bay sample, the only sample collected from a semi-enclosed bay, may reflect the interaction between coastal topography and recruitment.
- The number of SNPs revealed by RADseq was considerably lower than numbers reported for other marine taxa. SNP recovery was likely compromised by sub-optimal DNA quality and the lack of a reference genome. Global outlier analysis using approaches implemented in FDIST and BAYESCAN identified three consensus positive outlier SNPs. However, these outliers were not significant in outlier tests performed between pairs of samples, which are predicted to be more accurate (Robertson 1975). Furthermore, spatial patterns of allele frequency distribution at outlier loci did not permit linking of such variation with candidate environmental drivers of adaptation. However, the outlier analysis did highlight the potential of RADseq to identify particular SNPs that may provide enhanced resolution of stock structure (Limborg et al. 2012).

5.2. Implications for Fishery Management

- An important consideration when incorporating genetic data into fisheries management is that very limited exchange of migrants (1-5%) per generation is sufficient to obscure neutral marker genetic structuring (Slatkin 1993). While such low migration rates are generally sufficient for population connectivity over evolutionary timescales, they are a negligible force for replenishing depleted stocks over a timescale of interest to fisheries. As a result, weak or absent population structure may not reflect significant demographic connectivity among populations (Hauser & Carvalho 2008). However, the results of the biophysical modelling of larval dispersal indicating extensive spatial connectivity within the Irish Sea are compatible with the weak spatial structuring reflecting recurrent connectivity, at least for *H. gammarus*. Collectively these results indicate that in the absence of recognised stock structure, management enterprises for the Irish Sea must strive to encompass regional biocomplexity, as too localised an approach may not preserve spatially distinct source-sink links.
- Although often regarded as ephemeral and biologically insignificant, the chaotic genetic patchiness reported for *C. pagurus* has potentially important implications for community ecology (Booth & Brosnan 1995) and resource management (Larson & Julian 1999). As the patchiness is suggested to stem from stochastic recruitment processes, the optimal scale for marine reserve network design and fisheries management should vary in relation to spatial drivers of recruitment variability, thus highlighting the need to identify such drivers (see Section 5.3).
- An intriguing aspect of the results for *C. pagurus* is that the samples were composed of mixed cohorts of adults for which signatures of genetic structure would be expected to be diminished by larval and postlarval dispersal. These results indicate that: (i) stochastic recruitment processes may be even more extreme than revealed by our samples and/or; (ii) there may be higher levels of larval and postlarval cohesion than previously expected for this species. Both

features could mean that *C. pagurus* populations are more susceptible to environmental or anthropogenic impacts than previously thought.

5.3. Future Research

- Genetic analyses that focus on cohorts can yield important spatial and mechanistic insights into patterns of dispersal (Selkoe et al. 2006) and the processes underpinning chaotic genetic patchiness (e.g. Christie et al. 2010), and are recommended next steps for studies of population genetic structure in *H. gammarus* and *C. pagurus*. However, for both species, partitioning adults according to age is difficult and juveniles are highly elusive. An intriguing avenue for age structure analysis is the genotyping of larval samples. Species identification of early life history stages is currently employed as a management tool for a number of North Sea fisheries (McKeown et al. 2015) but could also be applied to within-species population level analysis. Such an approach would readily lend itself to a necessary genetic monitoring scheme to assess temporal changes in genetic variation (Ryman et al. 1995). If age structured analysis is not possible it is vital that some form of temporal surveys be performed as the loss of genetic variation may be more pronounced at various stages of exploitation (Ryman et al. 1995).
- Analysis of the neutral RADseq data suggests that even for increasingly powerful datasets, the lack of migration-drift equilibrium may complicate the interpretation of neutral genetic patterns in the context of spatial stock structure (as is often desirable for management). While future neutral genetic research should adopt a seascape approach (Selkoe et al. 2010) to identify determinants of recruitment patterns, markers under divergent selection are likely to be more informative stock diagnostic tools. Individual assignment and traceability of *C. harengus* in the Northeast Atlantic was achieved with 156 SNPs, of which 16 were identified as selective outliers. Bekkevold et al. (2015) is the first example of the development of a SNP-based tool for regional-scale genetic stock identification in fish outside Salmonids (Ackerman et al. 2011; Larson et al. 2012; Russello et al. 2012) and Atlantic cod (Bradbury et al. 2011; Pampoulie et al. 2011). Furthermore, their assignment method has much

improved levels of statistical power compared to previous studies utilising microsatellites and the currently applied morphological marker-based evaluations. Nielsen et al. (2012) was the first to demonstrate the applicability of SNP markers under selection to distinguish between *C. harengus* from the North Sea and the Northeast Atlantic. Together, these studies highlight the rapid rate of progression of genetic monitoring and stock identification in a weakly differentiated marine fish. Bekkevold et al. (2015) propose this approach as an adaptive tool to address a suite of biological, management, and forensic questions.

- While RADseq may reveal highly informative markers, due to the lack of a reference genome and thus difficulties in inferring functional variation, studies aimed at understanding local adaptation may be better performed by candidate gene genotyping (e.g. Larmuseau et al. 2009) in conjunction with seascape analysis.

6. REFERENCES

Ackerman M, Habicht C, Seeb L (2011) Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed stock analyses of sockeye salmon from Copper River, Alaska. Transactions of the American Fisheries Society, 140, 865-881.

Addison J, Hart M (2005) Spawning, copulation and inbreeding coefficients in marine invertebrates. Biology Letters, 1, 450-453.

Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences, 38, 1507-1514.

Anderson SC, Flemming JM, Watson R, Lotze HK (2011) Rapid global expansion of invertebrate fisheries: trends, drivers, and ecosystem effects. PLoS One, 6, e14735.

André C, Knutsen H (2010) Development of twelve novel microsatellite loci in the European lobster (*Homarus gammarus*). Conservation Genetics Resources, 2, 233-236.

André C, Larsson LC, Laikre L, Bekkevold D, Brigham J, Carvalho G, Dahlgren TG, Hutchinson W, Mariani S, Mudde K (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity, 106, 270-280.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. BMC Bioinformatics, 9, 323.

Atema J (1986) Review of sexual selection and chemical communication in the lobster, *Homarus americanus*. Canadian Journal of Fisheries and Aquatic Sciences, 43, 2283-2290.

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One, 3, e3376.

Banks SC, Piggott MP, Williamson JE, Bové U, Holbrook NJ, Beheregaray LB (2007) Oceanic variability and coastal topography shape genetic structure in a long-dispersing sea urchin. Ecology, 88, 3055-3064.

Bannister R, Addison J, Lovewell S (1994) Growth, movement, recapture rate and survival of hatchery-reared lobsters (*Homarus gammarus* (Linnaeus, 1758)) released into the wild on the English east coast. Crustaceana, 67, 156-172.

Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2000) Biogeography: a marine Wallace's line? Nature, 406, 692-693.

Batchelder HP (2006) Forward-in-Time-/Backward-in-Time-Trajectory (FITT/BITT) Modeling of Particles and Organisms in the Coastal Ocean. Journal of Atmospheric and Oceanic Technology, 23, 727-741.

Beacham TD, Jonsen K, Wallace C (2012) A comparison of stock and individual identification for Chinook salmon in British Columbia provided by microsatellites and single-nucleotide polymorphisms. Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science, 4, 1-22.

Beaumont MA (2005) Adaptation and speciation: what can F_{ST} tell us? Trends in Ecology & Evolution, 20, 435-440.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London Series B: Biological Sciences, 263, 1619-1626.

Becker BJ, Levin LA, Fodrie FJ, McMillan PA (2007) Complex larval connectivity patterns among marine invertebrate populations. Proceedings of the National Academy of Sciences, 104, 3267-3272.

Bekkevold D, Helyar SJ, Limborg MT, Nielsen EE, Hemmer-Hansen J, Clausen LA, Carvalho GR (2015) Gene-associated markers can assign origin in a weakly

structured fish, Atlantic herring. ICES Journal of Marine Science, doi: 10.1093/icesjms/fsu1247.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (1996-2004) GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Montpellier: Laboratoire Génome, Populations, Interactions, CNRS UMR 5171. Université de Montpellier II.

Bennett D, Brown C (1983) Crab (*Cancer pagurus*) migrations in the English Channel. Journal of the Marine Biological Association of the United Kingdom, 63, 371-398.

Bennett DB (1995) Factors in the life history of the edible crab (*Cancer pagurus* L.) that influence modelling and management. ICES Marine Science Symposia, 199, 89-98.

Blumberg AF, Mellor GL (1987) A description of a three-dimensional coastal ocean circulation model. Three-dimensional coastal ocean models. American Geophysics Union, New York.

Bohonak AJ (1999) Dispersal, gene flow, and population structure. Quarterly Review of Biology, 21-45.

Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P (2004) How to track and assess genotyping errors in population genetics studies. Molecular Ecology, 13, 3261-3273.

Booth DJ, Brosnan DM (1995) The role of recruitment dynamics in rocky shore and coral reef fish communities. Advances in Ecological Research, 26, 309-385.

Bradbury IR, Hubert S, Higgins B, Bowman S, Paterson IG, Snelgrove PV, Morris CJ, Gregory RS, Hardie DC, Borza T (2011) Evaluating SNP ascertainment bias and its impact on population assignment in Atlantic cod, *Gadus morhua*. Molecular Ecology Resources, 11, 218-225.

Bradbury IR, Laurel B, Snelgrove PV, Bentzen P, Campana SE (2008) Global patterns in marine dispersal estimates: the influence of geography, taxonomic category and life history. Proceedings of the Royal Society B: Biological Sciences, 275, 1803-1809.

Briggs JC (1974) Operation of zoogeographic barriers. Systematic Biology, 23, 248-256.

Brown C, Bennett D (1980) Population and catch structure of the edible crab (*Cancer pagurus*) in the English Channel. Journal du Conseil, 39, 88-100.

Cadrin SX, Kerr LA, Mariani S (2014) Stock identification methods: applications in fishery science. Second Edition. Academic Press.

Canino MF, O'Reilly PT, Hauser L, Bentzen P (2005) Genetic differentiation in walleye pollock (*Theragra chalcogramma*) in response to selection at the pantophysin (Pan I) locus. Canadian Journal of Fisheries and Aquatic Sciences, 62, 2519-2529.

Carvalho GR, Hauser L (1994) Molecular genetics and the stock concept in fisheries. Reviews in Fish Biology and Fisheries, 4, 326-350.

Cefas (2013) Shellfish News, Number 36, Autumn/Winter 2013. Accessed at: http://www.cefas.defra.gov.uk.

Chapuis M-P, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution, 24, 621-631.

Chasse J, Miller RJ (2010) Lobster larval transport in the southern Gulf of St. Lawrence. Fisheries Oceanography, 19, 319-338.

Christiansen ME (1969) Crustacea Decapoda Brachyura. In: Marine Invertebrates of Scandinavia. Scandinavian University Books, Universitetsforlaget Oslo.

Christie MR, Johnson DW, Stallings CD, Hixon MA (2010) Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. Molecular Ecology, 19, 1042-1057.

Cobb JS, Wahle RA (1994) Early life history and recruitment processes of clawed lobsters. Crustaceana, 1-25.

Conover D, Clarke L, Munch S, Wagner G (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. Journal of Fish Biology, 69, 21-47.

Corander J, Majander KK, Cheng L, Merilä J (2013) High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. Molecular Ecology, 22, 2931-2940.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics, 144, 2001-2014.

Coscia I, Robins PE, Porter JS, Malham SK, Ironside JE (2013) Modelled larval dispersal and measured gene flow: seascape genetics of the common cockle *Cerastoderma edule* in the southern Irish Sea. Conservation Genetics, 14, 451-466.

Coscia I, Vogiatzi E, Kotoulas G, Tsigenopoulos C, Mariani S (2012) Exploring neutral and adaptive processes in expanding populations of gilthead sea bream, *Sparus aurata* L., in the North-East Atlantic. Heredity, 108, 537-546.

Couceiro L, Barreiro R, Ruiz JM, Sotka EE (2007) Genetic isolation by distance among populations of the netted dog whelk *Nassarius reticulatus* (L.) along the European Atlantic coastline. Journal of Heredity, 98, 603-610.

Cowen R, Paris C, Srinivasan A (2006) Scaling of connectivity in marine populations. Science, 311, 522-527.

Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. Molecular Ecology Resources, 10, 556-557.

Crow JF, Kimura M (1970) An introduction to population genetics theory. Harper & Row, New York, NY.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST (2011) The variant call format and VCFtools. Bioinformatics, 27, 2156-2158.

Davies CE, Johnson AF, Wootton EC, Greenwood SJ, Clark KF, Vogan CL, Rowley AF (2014) Effects of population density and body size on disease ecology of the European lobster in a temperate marine conservation zone. ICES Journal of Marine Science, doi: 10.1093/icesjms/fsu1237.

Di Rienzo A, Peterson A, Garza J, Valdes A, Slatkin M, Freimer N (1994) Mutational processes of simple-sequence repeat loci in human populations. Proceedings of the National Academy of Sciences of the United States of America, 91, 3166-3170.

Domingues CP, Creer S, Taylor MI, Queiroga H, Carvalho GR (2010) Genetic structure of *Carcinus maenas* within its native range: larval dispersal and oceanographic variability. Marine Ecology Progress Series, 410, 111-123.

Duran S, Palacin C, Becerro MA, Turon X, Giribet G (2004) Genetic diversity and population structure of the commercially harvested sea urchin *Paracentrotus lividus* (Echinodermata, Echinoidea). Molecular Ecology, 13, 3317-3328.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4, 359-361.

Eaton D, Brown J, Addison J, Milligan S, Fernand L (2003) Edible crab (*Cancer pagurus*) larvae surveys off the east coast of England: implications for stock structure. Fisheries Research, 65, 191-199.

Edwards E (1966) Mating behaviour in the European edible crab (*Cancer pagurus* L.). Crustaceana, 10, 23-30.

Edwards E (1979) The Edible Crab and its Fishery in British Waters. Fishing News Books Ltd., Surrey, Great Britain.

El Mousadik A, Petit R (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theoretical and Applied Genetics, 92, 832-839.

Emerson KJ, Merz CR, Catchen JM, Hohenlohe PA, Cresko WA, Bradshaw WE, Holzapfel CM (2010) Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences, 107, 16196-16200.

Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics, 169, 1727-1738.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10, 564-567.

Factor JR (1995) Biology of the Lobster: *Homarus americanus*. Academic Press, New York and London.

Fahy E, Carroll J (2008) Two records of long migrations by Brown or Edible Crab (*Cancer pagurus* L.) from the Irish inshore of the Celtic Sea. The Irish Naturalists' Journal, 119-121.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164, 1567-1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes, 7, 574-578.

Faurby S, Barber PH (2012) Theoretical limits to the correlation between pelagic larval duration and population genetic structure. Molecular Ecology, 21, 3419-3432.

Fauvelot C, Planes S (2002) Understanding origins of present-day genetic structure in marine fish: biologically or historically driven patterns? Marine Biology, 141, 773-788.

Ferguson A (2002) Genetic diversity in the European lobster (*Homarus gammarus*): population structure and impacts of stock enhancement. Accessed at: http://www.qub.ac.uk/bb-old/prodohl/GEL/gel.html.

Fernand L, Nolan G, Raine R, Chambers C, Dye S, White M, Brown J (2006) The Irish coastal current: a seasonal jet-like circulation. Continental Shelf Research, 26, 1775-1793.

Fish JD, Fish S (2011) A student's guide to the seashore. Cambridge University Press.

Foll M, Gaggiotti OE (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, 180, 977-993.

Free E, Tyler P, Addison J (1992) Lobster (*Homarus gammarus*) fecundity and maturity in England and Wales. ICES Document CM.

Garza J, Williamson E (2001) Detection of reduction in population size using data from microsatellite loci. Molecular Ecology, 10, 305-318.

Gienapp P, Teplitsky C, Alho J, Mills J, Merilä J (2008) Climate change and evolution: disentangling environmental and genetic responses. Molecular Ecology, 17, 167-178.

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: coupling genetics with fine-scale physical oceanography. Ecology, 84, 2989-2998.

Glover KA, Dahle G, Jørstad KE (2011) Genetic identification of farmed and wild Atlantic cod, *Gadus morhua*, in coastal Norway. ICES Journal of Marine Science, 68, 901-910.

Goff LJ, Moon DA (1993) PCR amplification of nuclear and plastid genes from algal herbarium specimens and algal spores. Journal of Phycology, 29, 381-384.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software, 22, 1-19.

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of Heredity, 86, 485-486.

Graves J (1998) Molecular insights into the population structures of cosmopolitan marine fishes. Journal of Heredity, 89, 427-437.

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Lepoittevin C, Malausa T, Revardel E, Salin F (2011) Current trends in microsatellite genotyping. Molecular Ecology Resources, 11, 591-611.

Harding G, Drinkwater K, Hannah C, Pringle J, Prena J, Loder J, Pearre S, Vass W (2005) Larval lobster (*Homarus americanus*) distribution and drift in the vicinity of the Gulf of Maine offshore banks and their probable origins. Fisheries Oceanography, 14, 112-137.

Hart MW, Marko PB (2010) It's about time: divergence, demography, and the evolution of developmental modes in marine invertebrates. Integrative and Comparative Biology, doi:10.1093/icb/icq1068.

Hartnoll RG (1969) Mating in the Brachyura. Crustaceana, 16, 161-181.

Hartnoll RG, Bryant AD, Gould P (1993) Size distribution in spider crab populations: spatial and temporal variation. Journal of Crustacean Biology, 647-655.

Harvell C, Kim K, Burkholder J, Colwell R, Epstein PR, Grimes D, Hofmann E, Lipp E, Osterhaus A, Overstreet RM (1999) Emerging marine diseases: climate links and anthropogenic factors. Science, 285, 1505-1510.

Harvell CD, Mitchell CE, Ward JR, Altizer S, Dobson AP, Ostfeld RS, Samuel MD (2002) Climate warming and disease risks for terrestrial and marine biota. Science, 296, 2158-2162.

Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, Carvalho GR (2002) Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). Proceedings of the National Academy of Sciences, 99, 11742-11747.

Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries, 9, 333-362.

Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: Beaumont AR (ed) Genetics and evolution of aquatic organisms. Chapman & Hall, London. pp. 122-134.

Hedgecock D, Barber PH, Edmands S (2007) Genetic approaches to measuring connectivity. Oceanography, 20, 70-79.

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution, 59, 1633-1638.

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. Bulletin of Marine Science, 70, 273-290.

Hemmer-Hansen J, Nielsen E, Frydenberg J, Loeschcke V (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). Heredity, 99, 592-600.

Hoarau G, Boon E, Jongma DN, Ferber S, Palsson J, Van der Veer HW, Rijnsdorp AD, Stam WT, Olsen JL (2005) Low effective population size and evidence for inbreeding in an overexploited flatfish, plaice (*Pleuronectes platessa* L.). Proceedings of the Royal Society B: Biological Sciences, 272, 497-503.

Hoarau G, Piquet A-T, Van der Veer H, Rijnsdorp A, Stam W, Olsen J (2004) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: a comparison of resolving power between microsatellites and mitochondrial DNA data. Journal of Sea Research, 51, 183-190.

Hoarau G, Rijnsdorp A, Van der Veer H, Stam W, Olsen J (2002) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: microsatellites revealed large-scale spatial and temporal homogeneity. Molecular Ecology, 11, 1165-1176.

Hoegh-Guldberg O, Bruno JF (2010) The impact of climate change on the world's marine ecosystems. Science, 328, 1523-1528.

Hogan JD, Thiessen RJ, Heath DD (2010) Variability in connectivity indicated by chaotic genetic patchiness within and among populations of a marine fish. Marine Ecology Progress Series, 417, 263-289.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular Ecology Resources, 11, 117-122.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genetics, 6, e1000862.

Hoskin MG, Coleman RA, von Carlshausen E, Davis CM (2011) Variable population responses by large decapod crustaceans to the establishment of a temperate marine no-take zone. Canadian Journal of Fisheries and Aquatic Sciences, 68, 185-200.

Howard A, Bennett D (1979) The substrate preference and burrowing behaviour of juvenile lobsters (*Homarus gammarus* (L.)). Journal of Natural History, 13, 433-438.

Howard AE (1982) The distribution and behaviour of ovigerous edible crabs (*Cancer pagurus*), and consequent sampling bias. Journal du Conseil, 40, 259-261.

Howarth LM, Roberts CM, Thurstan RH, Stewart BD (2013) The unintended consequences of simplifying the sea: making the case for complexity. Fish and Fisheries, doi: 10.1111/faf.12041.

Hunter E, Eaton D, Stewart C, Lawler A, Smith MT (2013) Edible crabs "go west": migrations and incubation cycle of *Cancer pagurus* revealed by electronic tags. PloS One, 8, e63991.

Hutchings JA (2000) Collapse and recovery of marine fishes. Nature, 406, 882-885.

Hutchings JA, Reynolds JD (2004) Marine fish population collapses: consequences for recovery and extinction risk. Bioscience, 54, 297-309.

Hutchinson WF, van Oosterhout C, Rogers SI, Carvalho GR (2003) Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (*Gadus morhua*). Proceedings of the Royal Society of London B: Biological Sciences, 270, 2125-2132.

Imron BJ, Hale P, Degnan BM, Degnan SM (2007) Pleistocene isolation and recent gene flow in *Haliotis asinina*, an Indo-Pacific vetigastropod with limited dispersal capacity. Molecular Ecology, 16, 289-304.

Incze L, Xue H, Wolff N, Xu D, Wilson C, Steneck R, Wahle R, Lawton P, Pettigrew N, Chen Y (2010) Connectivity of lobster (*Homarus americanus*) populations in the coastal Gulf of Maine: part II. Coupled biophysical dynamics. Fisheries Oceanography, 19, 1-20.

Incze LS, Naimie CE (2000) Modelling the transport of lobster (*Homarus americanus*) larvae and postlarvae in the Gulf of Maine. Fisheries Oceanography, 9, 99-113.

Ingle R (1981) The larval and post-larval development of the edible crab, *Cancer pagurus* Linnaeus (Decapoda: Brachyura). Bulletin of the British Museum of Natural History, 40, 211-236.

James MK, Armsworth PR, Mason LB, Bode L (2002) The structure of reef fish metapopulations: modelling larval dispersal and retention patterns. Proceedings of the Royal Society of London Series B: Biological Sciences, 269, 2079-2086.

Jensen AC, Free EK, Collins KJ (1993) Lobster (*Homarus gammarus*) movement in the Poole Bay fishery, Dorset, UK. ICES Shellfish Committee, 1993/K 1949: 1991-1999. (Copenhagen).

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. BMC Genetics, 6, 13.

Johnson M, Black R (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. Marine Biology, 70, 157-164.

Johnson M, Hebbert D, Moran M (1993) Genetic analysis of populations of north-western Australian fish species. Marine and Freshwater Research, 44, 673-685.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24, 1403-1405.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27, 3070-3071.

Jones G, Almany G, Russ G, Sale P, Steneck R, Van Oppen M, Willis B (2009) Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. Coral Reefs, 28, 307-325.

Jost L (2008) G_{ST} and its relatives do not measure differentiation. Molecular Ecology, 17, 4015-4026.

Jørgensen HB, Hansen MM, Bekkevold D, Ruzzante DE, Loeschcke V (2005) Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. Molecular Ecology, 14, 3219-3234.

Jørstad KE, Farestveit E, Kelly E, Triantaphyllidis C (2005) Allozyme variation in European lobster (*Homarus gammarus*) throughout its distribution range. New Zealand Journal of Marine and Freshwater Research, 39, 515-526.

Jørstad KE, Prodöhl PA, Agnalt A-L, Hughes M, Apostolidis AP, Triantafyllidis A, Farestveit E, Kristiansen TS, Mercer J, Svåsand T (2004) Sub-arctic populations of

European lobster, *Homarus gammarus*, in northern Norway. Environmental Biology of Fishes, 69, 223-231.

Katz CH, Cobb JS, Spaulding M (1994) Larval behavior, hydrodynamic transport, and potential offshore-to-inshore recruitment in the American lobster *Homarus americanus*. Marine Ecology Progress Series, 103, 265-273.

Keenan K, McGinnity P, Cross TF, Crozier WW, Prodöhl PA (2013) diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution, 4, 782-788.

Kelly RP, Palumbi SR (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. PLoS One, 5, e8594.

Kennington W, Cadee S, Berry O, Groth D, Johnson M, Melville-Smith R (2013) Maintenance of genetic variation and panmixia in the commercially exploited western rock lobster (*Panulirus cygnus*). Conservation Genetics, 14, 115-124.

Kingsford MJ, Leis JM, Shanks A, Lindeman KC, Morgan SG, Pineda J (2002) Sensory environments, larval abilities and local self-recruitment. Bulletin of Marine Science, 70, 309-340.

Kinlan BP, Gaines SD, Lester SE (2005) Propagule dispersal and the scales of marine community process. Diversity and Distributions, 11, 139-148.

Knutsen H, Jorde P, André C, Stenseth N (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Molecular Ecology, 12, 385-394.

Knutsen H, Olsen EM, Jorde PE, Espeland SH, André C, Stenseth NC (2011) Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. Molecular Ecology, 20, 768-783.

Largier JL (2003) Considerations in estimating larval dispersal distances from oceanographic data. Ecological Applications, 13, 71-89.

Larmuseau MH, Raeymaekers JA, Ruddick KG, Van Houdt JK, Volckaert FA (2009) To see in different seas: spatial variation in the rhodopsin gene of the sand goby (*Pomatoschistus minutus*). Molecular Ecology, 18, 4227-4239.

Larsen PF, Nielsen EE, Williams TD, Hemmer-Hansen J, Chipman JK, Kruhøffer M, GRØNKJÆR P, George SG, DYRSKJØT L, Loeschcke V (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). Molecular Ecology, 16, 4674-4683.

Larson RJ, Julian R (1999) Spatial and temporal genetic patchiness in marine populations and their implications for fisheries management. California Cooperative Oceanic Fisheries Investigations Report, 40, 94-99.

Larson WA, Utter FM, Myers KW, Templin WD, Seeb JE, Guthrie III CM, Bugaev AV, Seeb LW, Moran P (2012) Single-nucleotide polymorphisms reveal distribution and migration of Chinook salmon (*Oncorhynchus tshawytscha*) in the Bering Sea and North Pacific Ocean. Canadian Journal of Fisheries and Aquatic Sciences, 70, 128-141.

Latch EK, Dharmarajan G, Glaubitz JC, Rhodes Jr OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. Conservation Genetics, 7, 295-302.

Latrouite D, Le Foll D (1989) Données sur les migrations des crabes tourteau *Cancer pagurus* et les araignées de mer *Maja squinado*. Océanis, 15, 133-142.

Laurans M, Fifas S, Demaneche S, Brérette S, Debec O (2009) Modelling seasonal and annual variation in size at functional maturity in the European lobster (*Homarus gammarus*) from self-sampling data. ICES Journal of Marine Science, 66, 1892-1898.

Levin LA (2006) Recent progress in understanding larval dispersal: new directions and digressions. Integrative and Comparative Biology, 46, 282-297.

Li G, Hedgecock D (1998) Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. Canadian Journal of Fisheries and Aquatic Sciences, 55, 1025-1033.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078-2079.

Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research, 18, 1851-1858.

Limborg MT, Helyar SJ, de Bruyn M, Taylor MI, Nielsen EE, Ogden R, Carvalho GR, Bekkevold D (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). Molecular Ecology, 21, 3686-3703.

Lizárraga-Cubedo H, Tuck I, Bailey N, Pierce G, Kinnear J (2003) Comparisons of size at maturity and fecundity of two Scottish populations of the European lobster, *Homarus gammarus*. Fisheries Research, 65, 137-152.

Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? Molecular Ecology, 19, 3038-3051.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. Journal of Heredity, 89, 238-247.

Luikart G, Cornuet JM (1998) Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conservation Biology, 12, 228-237.

MacKenzie BR (1988) Assessment of temperature effects on interrelationships between stage durations, mortality, and growth in laboratory-reared *Homarus americanus* Milne Edwards larvae. Journal of Experimental Marine Biology and Ecology, 116, 87-98.

MacKenzie BR, Payne MR, Boje J, Høyer JL, Siegstad H (2014) A cascade of warming impacts brings bluefin tuna to Greenland waters. Global Change Biology, 20, 2484-2491.

Malham S (2010) The SUSFISH Project. Shellfish News (CEFAS), Number 30, Autumn/Winter 2010. Accessed at: http://www.cefas.defra.gov.uk.

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Research, 27, 209-220.

Mardis ER (2008) Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics, 9, 387-402.

Marine Management Organisation (2014) UK Sea Fisheries Statistics 2013 (ed. Radford L). Accessed at: www.gov.uk/government/organisations/marine-management-organisation.

Marko P (2004) 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. Molecular Ecology, 13, 597-611.

McKeown NJ, Robin J-P, Shaw PW (2015) Species-specific PCR-RFLP for identification of early life history stages of squid and other applications to fisheries research. Fisheries Research, 167, 207-209.

McKeown NJ, Shaw PW (2008a) Polymorphic nuclear microsatellite loci for studies of brown crab, *Cancer pagurus* L. Molecular Ecology Resources, 8, 653-655.

McKeown NJ, Shaw PW (2008b) Single paternity within broods of the brown crab *Cancer pagurus*: a highly fecund species with long-term sperm storage. Marine Ecology Progress Series, 368, 209-215.

Mellor GL, Yamada T (1982) Development of a turbulence closure model for geophysical fluid problems. Reviews of Geophysics, 20, 851-875.

Metaxas A, Saunders M (2009) Quantifying the "bio-" components in biophysical models of larval transport in marine benthic invertebrates: advances and pitfalls. The Biological Bulletin, 216, 257-272.

Milano I, Babbucci M, Cariani A, Atanassova M, Bekkevold D, Carvalho GR, Espiñeira M, Fiorentino F, Garofalo G, Geffen AJ (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). Molecular Ecology, 23, 118-135.

Miller LM, Kapuscinski AR (1997) Historical analysis of genetic variation reveals low effective population size in a northern pike (*Esox lucius*) population. Genetics, 147, 1249-1258.

Moland E, Olsen EM, Knutsen H, Garrigou P, Espeland SH, Kleiven AR, André C, Knutsen JA (2013) Lobster and cod benefit from small-scale northern marine protected areas: inference from an empirical before–after control-impact study. Proceedings of the Royal Society B: Biological Sciences, 280, 20122679.

Moland E, Olsen EM, Knutsen H, Knutsen JA, Enersen SE, André C, Stenseth NC (2011) Activity patterns of wild European lobster *Homarus gammarus* in coastal marine reserves: implications for future reserve design. Marine Ecology Progress Series, 429, 197-207.

Molfese C, Beare D, Hall-Spencer JM (2014) Overfishing and the replacement of demersal finfish by shellfish: an example from the English Channel. PloS One, 9, e101506.

Moran B (2009) Population genetic structure of brown crab (*Cancer pagurus*) in Irish waters. MSc Thesis, Galway-Mayo Institute of Technology.

Mulley J, Latter B (1981) Geographic differentiation of tropical Australian penaeid prawn populations. Marine and Freshwater Research, 32, 897-906.

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. Molecular Ecology Resources, 11, 184-194.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, 89, 583-590.

Neill SP, Litt EJ, Couch SJ, Davies AG (2009) The impact of tidal stream turbines on large-scale sediment dynamics. Renewable Energy, 34, 2803-2812.

Neill SP, Scourse JD, Uehara K (2010) Evolution of bed shear stress distribution over the northwest European shelf seas during the last 12,000 years. Ocean Dynamics, 60, 1139-1156.

Nichols J, Lovewell S (1987) Lobster larvae (*Homarus gammarus* L.) investigations in Bridlington Bay. Can quantitative sampling be confined to the neuston layer? Journal of Natural History, 21, 825-841.

Nielsen EE, Cariani A, Mac Aoidh E, Maes GE, Milano I, Ogden R, Taylor M, Hemmer-Hansen J, Babbucci M, Bargelloni L (2012) Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. Nature Communications, 3, 851.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. Molecular Ecology, 18, 3128-3150.

Nielsen EE, Kenchington E (2001) A new approach to prioritizing marine fish and shellfish populations for conservation. Fish and Fisheries, 2, 328-343.

North E, Schlag Z, Hood R, Li M, Zhong L, Gross T, Kennedy V (2008) Vertical swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle-tracking and hydrodynamic model of Chesapeake Bay. Marine Ecology Progress Series, 359, 99-115.

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. Genetical Research, 22, 201-204.

Olsen EM, Heino M, Lilly GR, Morgan MJ, Brattey J, Ernande B, Dieckmann U (2004) Maturation trends indicative of rapid evolution preceded the collapse of northern cod. Nature, 428, 932-935.

Øresland V, Ulmestrand M (2013) European lobster subpopulations from limited adult movements and larval retention. ICES Journal of Marine Science, 70, 532-539.

Ouellet P, Allard J-P (2002) Seasonal and interannual variability in larval lobster *Homarus americanus* size, growth and condition in the Magdalen Islands, southern Gulf of St. Lawrence. Marine Ecology Progress Series, 230, 241-251.

Palero F, Abello P, Macpherson E, Beaumont M, Pascual M (2011) Effect of oceanographic barriers and overfishing on the population genetic structure of the

European spiny lobster (*Palinurus elephas*). Biological Journal of the Linnean Society, 104, 407-418.

Palumbi SR (1992) Marine speciation on a small planet. Trends in Ecology & Evolution, 7, 114-118.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics, 547-572.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications, 13, 146-158.

Pampoulie C, Daníelsdóttir AK, Storr-Paulsen M, Hovgård H, Hjörleifsson E, Steinarsson BÆ (2011) Neutral and nonneutral genetic markers revealed the presence of inshore and offshore stock components of Atlantic cod in Greenland waters. Transactions of the American Fisheries Society, 140, 307-319.

Pampoulie C, Ruzzante DE, Chosson V, Jörundsdóttir TD, Taylor L, Thorsteinsson V, Daníelsdóttir AK, Marteinsdóttir G (2006) The genetic structure of Atlantic cod (*Gadus morhua*) around Iceland: insight from microsatellites, the Pan I locus, and tagging experiments. Canadian Journal of Fisheries and Aquatic Sciences, 63, 2660-2674.

Park SDE (2001) Trypanotolerance in West African cattle and the population genetic effects of selection. PhD Thesis, University of Dublin.

Paschou P, Ziv E, Burchard EG, Choudhry S, Rodriguez-Cintron W, Mahoney MW, Drineas P (2007) PCA-correlated SNPs for structure identification in worldwide human populations. PLoS Genetics, 3, e160.

Pawson MG (1995) Biogeographical identification of English Channel fish and shellfish stocks. Fisheries Resources Technical Report Number 99. Lowestoft: Ministry of Agriculture, Fisheries and Food.

Pearson J (1908) *Cancer*. Liverpool Marine Biology Committee Memoirs, 16.

Pechenik JA (1999) On the advantages and disadvantages of larval stages in benthic marine invertebrate life cycles. Marine Ecology Progress Series, 177, 269-297.

Peijnenburg K, Fauvelot C, Breeuwer J, Menken S (2006) Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. Molecular Ecology, 15, 3319-3338.

Piry S, Luikart G, Cornuet J-M (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. Journal of Heredity, 90, 502-503.

Planes S, Lenfant P (2002) Temporal change in the genetic structure between and within cohorts of a marine fish, *Diplodus sargus*, induced by a large variance in individual reproductive success. Molecular Ecology, 11, 1515-1524.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Prodöhl P, Jørstad K, Triantafyllidis A, Katsares V, Triantaphyllidis C (2007) European lobster - *Homarus gammarus*. Genimpact Final Scientific Report, 91-98. Accessed at: http://www.imr.no/genimpact/__data/page/7650/european_lobster.pdf.

Pujolar J, Maes G, Volckaert F (2006) Genetic patchiness among recruits in the European eel *Anguilla anguilla*. Marine Ecology Progress Series, 307, 209-217.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity, 86, 248-249.

Reitzel A, Herrera S, Layden M, Martindale M, Shank T (2013) Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Molecular Ecology, 22, 2953-2970.

Rice WR (1989) Analyzing tables of statistical tests. Evolution, 223-225.

Riginos C, Douglas KE, Jin Y, Shanahan DF, Treml EA (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. Ecography, 34, 566-575.

Rioux Paquette S (2011) PopGenKit: useful functions for (batch) file conversion and data resampling in microsatellite datasets. R package version 1.0. Accessed at: http://cran.r-project.org/web/packages/PopGenKit/index.html.

Robertson A (1975) Gene frequency distributions as a test of selective neutrality. Genetics, 81, 775-785.

Robins P, Neill S, Giménez L (2012) A numerical study of marine larval dispersal in the presence of an axial convergent front. Estuarine, Coastal and Shelf Science, 100, 172-185.

Robins PE, Neill SP, Gimenez L, Jenkins SR, Malham SK (2013) Physical and biological controls on larval dispersal and connectivity in a highly energetic shelf sea. Limnology and Oceanography, 58, 489-504.

Robinson I (1979) The tidal dynamics of the Irish and Celtic Seas. Geophysical Journal International, 56, 159-197.

Ross PM, Hogg ID, Pilditch CA, Lundquist CJ (2009) Phylogeography of New Zealand's coastal benthos. New Zealand Journal of Marine and Freshwater Research, 43, 1009-1027.

Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources, 8, 103-106.

Rowe H, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. Molecular Ecology, 20, 3499-3502.

Russello MA, Kirk SL, Frazer KK, Askey PJ (2012) Detection of outlier loci and their utility for fisheries management. Evolutionary Applications, 5, 39-52.

Ruzzante DE, Taggart CT, Cook D (1996) Spatial and temporal variation in the genetic composition of a larval cod (*Gadus morhua*) aggregation: cohort contribution and genetic stability. Canadian Journal of Fisheries and Aquatic Sciences, 53, 2695-2705.

Ryman N, Lagercrantz U, Andersson L, Chakraborty R, Rosenberg R (1984) Lack of correspondence between genetic and morphologic variability patterns in Atlantic herring (*Clupea harengus*). Heredity, 53, 687-704.

Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Molecular Ecology Notes, 6, 600-602.

Ryman N, Utter F, Laikre L (1995) Protection of intraspecific biodiversity of exploited fishes. Reviews in Fish Biology and Fisheries, 5, 417-446.

SEAFISH (2013) Responsible sourcing guide: crabs and lobsters. Version 4. Accessed at: http://www.seafish.org/media/publications/SeafishResponsibleSourcingGuide_CrabsLobsters_201309.pdf.

Selkoe KA, Gaines SD, Caselle JE, Warner RR (2006) Current shifts and kin aggregation explain genetic patchiness in fish recruits. Ecology, 87, 3082-3094.

Selkoe KA, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Marine Ecology Progress Series, 436, 291-305.

Selkoe KA, Watson JR, White C, Ben-Horin T, Iacchei M, Mitarai S, Siegel DA, Gaines SD, Toonen RJ (2010) Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. Molecular Ecology, 19, 3708-3726.

Shaklee JB, Bentzen P (1998) Genetic identification of stocks of marine fish and shellfish. Bulletin of Marine Science, 62, 589-621.

Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. The Biological Bulletin, 216, 373-385.

Shaw P, Hendrickson L, McKeown N, Stonier T, Naud M, Sauer W (2010) Discrete spawning aggregations of loliginid squid do not represent genetically distinct populations. Marine Ecology Progress Series, 408, 117-127.

Shaw P, Turan C, Wright JM, O'connell M, Carvalho G (1999) Microsatellite DNA analysis of population structure in Atlantic herring (*Clupea harengus*), with direct comparison to allozyme and mtDNA RFLP analyses. Heredity, 83, 490-499.

Shaw PW (2003) Population genetics of shellfish in British waters. Final report for DEFRA, Project MF0226.

Sheehy M, Bannister R, Wickins J, Shelton P (1999) New perspectives on the growth and longevity of the European lobster (*Homarus gammarus*). Canadian Journal of Fisheries and Aquatic Sciences, 56, 1904-1915.

Siegel D, Kinlan B, Gaylord B, Gaines S (2003) Lagrangian descriptions of marine larval dispersion. Marine Ecology Progress Series, 260, 83-96.

Silberman J, Walsh P (1994) Population genetics of the spiny lobster, *Panulirus argus*. Bulletin of Marine Science, 54, 1084.

Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. Evolution, 47, 264-279.

Slavov GT, Nipper R, Robson P, Farrar K, Allison GG, Bosch M, Clifton-Brown JC, Donnison IS, Jensen E (2014) Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. New Phytologist, 201, 1227-1239.

Smith I, Jensen A, Collins K, Mattey E (2001) Movement of wild European lobsters *Homarus gammarus* in natural habitat. Marine Ecology Progress Series, 222, 177-186.

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology, 35, 627-632.

Sotelo G, Morán P, Fernández L, Posada D (2008) Genetic variation of the spiny spider crab *Maja brachydactyla* in the northeastern Atlantic. Marine Ecology Progress Series, 362, 211-223.

Sotelo G, Posada D, Morán P (2009) Low-mitochondrial diversity and lack of structure in the velvet swimming crab *Necora puber* along the Galician coast. Marine Biology, 156, 1039-1048.

Stamatis C, Triantafyllidis A, Moutou K, Mamuris Z (2004) Mitochondrial DNA variation in Northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus*. Molecular Ecology, 13, 1377-1390.

Stamatis C, Triantafyllidis A, Moutou KA, Mamuris Z (2006) Allozymic variation in Northeast Atlantic and Mediterranean populations of Norway lobster, *Nephrops norvegicus*. ICES Journal of Marine Science, 63, 875-882.

Swearer SE, Caselle JE, Lea DW, Warner RR (1999) Larval retention and recruitment in an island population of a coral-reef fish. Nature, 402, 799-802.

Tallack SM (2007a) Size–fecundity relationships for *Cancer pagurus* and *Necora puber* in the Shetland Islands, Scotland: how is reproductive capacity facilitated? Journal of the Marine Biological Association of the United Kingdom, 87, 507-515.

Tallack SM (2007b) The reproductive cycle and size at maturity observed in *Cancer pagurus* in the Shetland Islands, Scotland. Journal of the Marine Biological Association of the United Kingdom, 87, 1181-1189.

Taylor G (1919) Tidal friction in the Irish Sea. Philosophical Transactions of the Royal Society of London Series A, Containing Papers of a Mathematical or Physical Character, 220, 1-33.

Taylor MS, Hellberg ME (2003) Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. Science, 299, 107-109.

Teske P, Papadopoulos I, Zardi G, McQuaid C, Edkins M, Griffiths C, Barker N (2007) Implications of life history for genetic structure and migration rates of southern African coastal invertebrates: planktonic, abbreviated and direct development. Marine Biology, 152, 697-711.

Thorpe J, Solé-Cava A, Watts P (2000) Exploited marine invertebrates: genetics and fisheries. Hydrobiologia, 420, 165-184.

Todd CD, Lambert WJ, Thorpe JP (1998) The genetic structure of intertidal populations of two species of nudibranch molluscs with planktotrophic and pelagic lecithotrophic larval stages: are pelagic larvae “for” dispersal? Journal of Experimental Marine Biology and Ecology, 228, 1-28.

Toonen RJ, Grosberg RK (2011) Causes of chaos: spatial and temporal genetic heterogeneity in the intertidal anomuran crab *Petrolisthes cinctipes*. In: Koenemann S, Held C, Schubart C (eds) Crustacean Issues: Phylogeography and Population Genetics in Crustacea. CRC Press, Boca Raton, Florida. pp. 75-107.

Triantafyllidis A, Apostolidis A, Katsares V, Kelly E, Mercer J, Hughes M, Jørstad K, Tsolou A, Hynes R, Triantaphyllidis C (2005) Mitochondrial DNA variation in the European lobster (*Homarus gammarus*) throughout the range. Marine Biology, 146, 223-235.

Tully O, Roantree V, Robinson M (2001) Maturity, fecundity and reproductive potential of the European lobster (*Homarus gammarus*) in Ireland. Journal of the Marine Biological Association of the United Kingdom, 81, 61-68.

Turner TF, Wares JP, Gold JR (2002) Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). Genetics, 162, 1329-1339.

Ungfors A (2007) Sexual maturity of the edible crab (*Cancer pagurus*) in the Skagerrak and the Kattegat, based on reproductive and morphometric characters. ICES Journal of Marine Science, 64, 318-327.

Ungfors A, Hallbäck H, Nilsson PG (2007) Movement of adult edible crab (*Cancer pagurus* L.) at the Swedish West Coast by mark-recapture and acoustic tracking. Fisheries Research, 84, 345-357.

Ungfors A, McKeown NJ, Shaw PW, André C (2009) Lack of spatial genetic variation in the edible crab (*Cancer pagurus*) in the Kattegat–Skagerrak area. ICES Journal of Marine Science, 66, 462-469.

Utter F (1991) Biochemical genetics and fishery management: an historical perspective. Journal of Fish Biology, 39, 1-20.

Vasemägi A, Primmer C (2005) Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. Molecular Ecology, 14, 3623-3642.

Waddy S, Aiken D (1986) Multiple fertilization and consecutive spawning in large American lobsters, *Homarus americanus*. Canadian Journal of Fisheries and Aquatic Sciences, 43, 2291-2294.

Wahle RA, Castro KM, Tully O, Cobb JS (2013) *Homarus*. In: Phillips BF (ed) Lobsters: Biology, Management, Aquaculture and Fisheries, Second Edition. John Wiley & Sons, Ltd., Somerset, NJ, USA. pp. 221-258.

Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques, 10, 506-513.

Waples R (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity, 89, 438-450.

Waples RS, Do C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources, 8, 753-756.

Ward R (2000) Genetics in fisheries management. Hydrobiologia, 420, 191-201.

Ward R, Woodwark M, Skibinski D (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of Fish Biology, 44, 213-232.

Weber L, Hartnoll R, Thorpe J (2000) Genetic divergence and larval dispersal in two spider crabs (Crustacea: Decapoda). Hydrobiologia, 420, 211-219.

Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. Marine Ecology Progress Series, 393, 1-12.

Weetman D, Ruggiero A, Mariani S, Shaw P, Lawler A, Hauser L (2007) Hierarchical population genetic structure in the commercially exploited shrimp *Crangon crangon* identified by AFLP analysis. Marine Biology, 151, 565-575.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution, 1358-1370.

Weiss M, Thatje S, Heilmayer O, Anger K, Brey T, Keller M (2009) Influence of temperature on the larval development of the edible crab, *Cancer pagurus*. Journal of the Marine Biological Association of the United Kingdom, 89, 753-759.

Welsh Government (2014) Crustacean Fishery Consultation - Summarised Proposals. At: gov.wales/consultations/environmentandcountryside/crustaceans-legislation-review-phase-1.

White C, Selkoe KA, Watson J, Siegel DA, Zacherl DC, Toonen RJ (2010) Ocean currents help explain population genetic structure. Proceedings of the Royal Society B: Biological Sciences, 1685-1694.

Whitlock MC (2011) G'_{ST} and D do not replace F_{ST}. Molecular Ecology, 20, 1083-1091.

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics, 163, 1177-1191.

Woll AK, van der Meeren GI, Fossen I (2006) Spatial variation in abundance and catch composition of *Cancer pagurus* in Norwegian waters: biological reasoning and implications for assessment. ICES Journal of Marine Science, 63, 421-433.

Wootton EC, Woolmer AP, Vogan CL, Pope EC, Hamilton KM, Rowley AF (2012) Increased disease calls for a cost-benefits review of marine reserves. PloS One, 7, e51615.

Wright S (1943) Isolation by distance. Genetics, 28, 114-138.

Xue H, Incze L, Xu D, Wolff N, Pettigrew N (2008) Connectivity of lobster populations in the coastal Gulf of Maine: Part I: Circulation and larval transport potential. Ecological Modelling, 210, 193-211.

7. APPENDICES

Appendix 1. Linkage disequilibrium in *Homarus gammarus* in the Irish Sea.

Locus pair			Chi^2	df	P-value
HGD111	&	HGD106	13.380	18	0.769
HGD111	&	HGC131b	11.566	18	0.869
HGD106	**&**	**HGC131b**	**31.360**	**18**	**0.026***
HGD111	&	HGC129	17.139	18	0.514
HGD106	&	HGC129	12.257	18	0.834
HGC131b	&	HGC129	11.167	18	0.887
HGD111	&	HGC120	28.385	18	0.056
HGD106	&	HGC120	4.137	18	1.000
HGC131b	&	HGC120	15.031	18	0.660
HGC129	&	HGC120	10.245	18	0.924
HGD111	&	HGB6	17.498	18	0.489
HGD106	&	HGB6	15.279	18	0.643
HGC131b	&	HGB6	20.897	18	0.285
HGC129	&	HGB6	20.216	18	0.321
HGC120	&	HGB6	12.695	18	0.809
HGD111	&	HGC118	15.484	18	0.629
HGD106	&	HGC118	16.323	18	0.570
HGC131b	&	HGC118	16.380	18	0.566
HGC129	**&**	**HGC118**	**31.960**	**18**	**0.022***
HGC120	&	HGC118	23.061	18	0.188
HGB6	&	HGC118	17.833	18	0.467
HGD111	&	HGC111	11.610	18	0.867
HGD106	&	HGC111	18.494	18	0.424
HGC131b	&	HGC111	15.377	18	0.636
HGC129	&	HGC111	27.013	18	0.079
HGC120	&	HGC111	7.458	18	0.986
HGB6	&	HGC111	14.084	18	0.724
HGC118	&	HGC111	22.411	18	0.214
HGD111	**&**	**HGC103**	**29.772**	**18**	**0.040***
HGD106	&	HGC103	13.245	18	0.777
HGC131b	&	HGC103	14.177	18	0.717
HGC129	&	HGC103	13.400	18	0.767
HGC120	&	HGC103	21.284	18	0.265

Locus pair			Chi^2	df	P-value
HGB6	&	HGC103	7.548	18	0.985
HGC118	&	HGC103	24.477	18	0.140
HGC111	&	HGC103	9.229	18	0.954
HGD111	&	HGC6	22.008	18	0.232
HGD106	&	HGC6	24.043	18	0.154
HGC131b	&	HGC6	9.506	18	0.947
HGC129	&	HGC6	20.076	18	0.329
HGC120	&	HGC6	20.375	18	0.312
HGB6	&	HGC6	16.114	18	0.585
HGC118	&	HGC6	14.511	18	0.695
HGC111	&	HGC6	18.317	18	0.435
HGC103	&	HGC6	21.936	18	0.235
HGD111	&	HGB4	11.295	18	0.881
HGD106	&	HGB4	17.679	18	0.477
HGC131b	&	HGB4	11.898	18	0.852
HGC129	&	HGB4	14.496	18	0.696
HGC120	&	HGB4	18.265	18	0.438
HGB6	&	HGB4	24.017	18	0.154
HGC118	&	HGB4	9.630	18	0.943
HGC111	&	HGB4	13.439	18	0.765
HGC103	&	HGB4	13.651	18	0.752
HGC6	&	HGB4	15.144	18	0.652
HGD111	&	HGA8	9.248	18	0.954
HGD106	&	HGA8	18.252	18	0.439
HGC131b	&	HGA8	24.483	18	0.140
HGC129	&	HGA8	7.595	18	0.984
HGC120	&	HGA8	3.181	18	1.000
HGB6	&	HGA8	14.240	18	0.713
HGC118	&	HGA8	11.456	18	0.874
HGC111	&	HGA8	14.364	18	0.705
HGC103	&	HGA8	16.009	18	0.592
HGC6	&	HGA8	16.353	18	0.568
HGB4	&	HGA8	13.220	18	0.778

Appendix 2. F_{ST} values for *Homarus gammarus* in the Irish Sea before (upper diagonal) and after (lower diagonal) ENA correction.

	LNTZ	LICZ	WF	CB	ND	WEX	SW	NW	DEV
LNTZ		-0.002	0.002	0.000	-0.002	-0.001	0.001	0.000	0.000
LICZ	-0.001		-0.002	0.001	0.000	0.001	-0.002	-0.002	-0.001
WF	0.003	-0.001		0.002	0.004	0.004	0.000	0.003	0.000
CB	0.001	0.001	0.003		-0.001	0.000	-0.001	-0.004	-0.001
ND	-0.001	0.000	0.004	0.000		0.003	0.000	-0.003	-0.002
WEX	0.001	0.001	0.005	0.000	0.003		0.001	0.001	0.003
SW	0.002	-0.002	0.001	-0.001	0.000	0.001		-0.002	-0.003
NW	0.000	-0.002	0.004	-0.004	-0.003	0.001	-0.002		-0.004
DEV	0.001	-0.001	0.001	-0.001	-0.002	0.003	-0.003	-0.004	

Appendix 3. IBD matrix comprising of genetic distance (F_{ST} - upper diagonal) and the corresponding geographic distance (km - lower diagonal) for *Homarus gammarus* in the Irish Sea.

IBD	LNTZ	LICZ	WF	CB	ND	WEX	SW	NW	DEV
LNTZ		-0.002	0.002	0.000	-0.002	-0.001	0.001	0.000	0.000
LICZ	3.923		-0.002	0.001	0.000	0.001	-0.002	-0.002	-0.001
WF	193.893	193.547		0.002	0.004	0.004	0.000	0.004	0.000
CB	204.524	205.874	200.043		-0.001	0.000	-0.001	-0.004	-0.001
ND	281.044	282.033	197.709	176.299		0.003	0.000	-0.003	-0.002
WEX	163.081	163.622	48.425	154.400	149.793		0.001	0.001	0.003
SW	54.364	57.199	213.205	217.111	291.605	182.556		-0.002	-0.003
NW	208.233	209.163	173.107	64.641	113.436	126.756	215.977		-0.004
DEV	37.237	38.701	221.631	224.358	303.400	187.909	35.762	232.726	

Appendix 4. The matrix of inferred (posterior mean) migration rates and the standard deviation of the marginal posterior distribution for each estimate. m[i][j] is the fraction of individuals in population [i] that are migrants derived from population [j] (per generation), with 0 = LNTZ, 1 = SW, 2 = DEV, and 3 = WF & WEX.

Simulation 1			
m[0][0]: 0.6736(0.0067)	m[0][1]: 0.0142(0.0101)	**m[0][2]: 0.2762(0.0282)**	m[0][3]: 0.0361(0.0263)
m[1][0]: 0.0072(0.0071)	m[1][1]: 0.6731(0.0064)	**m[1][2]: 0.3052(0.0158)**	m[1][3]: 0.0144(0.0130)
m[2][0]: 0.0120(0.0115)	m[2][1]: 0.0159(0.0155)	m[2][2]: 0.8951(0.0397)	m[2][3]: 0.0770(0.0373)
m[3][0]: 0.0064(0.0062)	m[3][1]: 0.0062(0.0059)	**m[3][2]: 0.3039(0.0157)**	m[3][3]: 0.6835(0.0138)

Simulation 2			
m[0][0]: 0.6735(0.0067)	m[0][1]: 0.0143(0.0100)	m[0][2]: 0.0094(0.0092)	**m[0][3]: 0.3028(0.0142)**
m[1][0]: 0.0070(0.0069)	m[1][1]: 0.6734(0.0067)	m[1][2]: 0.0080(0.0081)	**m[1][3]: 0.3115(0.0122)**
m[2][0]: 0.0072(0.0069)	m[2][1]: 0.0077(0.0077)	m[2][2]: 0.6741(0.0073)	**m[2][3]: 0.3110(0.0124)**
m[3][0]: 0.0092(0.0082)	m[3][1]: 0.0103(0.0090)	m[3][2]: 0.0121(0.0122)	m[3][3]: 0.9685(0.0172)

Appendix 5. Linkage disequilibrium in *Cancer pagurus* in the Irish Sea.

Locus pair			Chi^2	df	P-value
Cpag 4	&	Cpag 1B9	15.280	22	0.850
Cpag 4	&	Cpag 1C8	2.908	18	1.000
Cpag 1B9	&	Cpag 1C8	12.481	22	0.947
Cpag 4	&	Cpag 2A5B	16.029	22	0.814
Cpag 1B9	&	Cpag 2A5B	18.575	26	0.854
Cpag 1C8	&	Cpag 2A5B	17.988	22	0.707
Cpag 4	&	Cpag 3A2	17.641	22	0.727
Cpag 1B9	&	Cpag 3A2	20.172	26	0.783
Cpag 1C8	&	Cpag 3A2	13.109	22	0.930
Cpag 2A5B	&	Cpag 3A2	22.659	26	0.652
Cpag 4	&	Cpag 3D7	13.601	22	0.915
Cpag 1B9	&	Cpag 3D7	29.259	26	0.299
Cpag 1C8	&	Cpag 3D7	9.944	22	0.987
Cpag 2A5B	&	Cpag 3D7	15.613	26	0.945
Cpag 3A2	&	Cpag 3D7	21.902	26	0.694
Cpag 4	&	Cpag 4C1	13.820	22	0.908
Cpag 1B9	&	Cpag 4C1	18.404	26	0.861
Cpag 1C8	&	Cpag 4C1	20.469	22	0.554
Cpag 2A5B	&	Cpag 4C1	21.813	26	0.699
Cpag 3A2	&	Cpag 4C1	25.606	26	0.485
Cpag 3D7	&	Cpag 4C1	27.864	26	0.365
Cpag 4	&	Cpag 5D8	5.676	22	1.000
Cpag 1B9	&	Cpag 5D8	15.160	24	0.916
Cpag 1C8	&	Cpag 5D8	11.451	20	0.934
Cpag 2A5B	&	Cpag 5D8	18.603	24	0.773
Cpag 3A2	&	Cpag 5D8	19.673	24	0.715
Cpag 3D7	&	Cpag 5D8	5.911	24	1.000
Cpag 4C1	&	Cpag 5D8	18.092	24	0.799

Locus pair			Chi^2	df	P-value
Cpag 4	&	Cpag 6C4B	4.922	22	1.000
Cpag 1B9	&	Cpag 6C4B	20.378	26	0.773
Cpag 1C8	&	Cpag 6C4B	23.235	22	0.389
Cpag 2A5B	&	Cpag 6C4B	16.321	26	0.928
Cpag 3A2	&	Cpag 6C4B	17.406	26	0.896
Cpag 3D7	&	Cpag 6C4B	27.757	26	0.371
Cpag 4C1	&	Cpag 6C4B	22.201	26	0.678
Cpag 5D8	&	Cpag 6C4B	21.504	24	0.609
Cpag 4	&	Cpag 15	18.789	22	0.658
Cpag 1B9	&	Cpag 15	23.767	26	0.589
Cpag 1C8	&	Cpag 15	10.651	22	0.979
Cpag 2A5B	&	Cpag 15	31.948	26	0.195
Cpag 3A2	&	Cpag 15	25.082	26	0.514
Cpag 3D7	&	Cpag 15	29.859	26	0.274
Cpag 4C1	&	Cpag 15	23.865	26	0.584
Cpag 5D8	&	Cpag 15	14.637	24	0.931
Cpag 6C4B	&	Cpag 15	15.179	26	0.954
Cpag 4	&	Cpag 38	1.794	22	1.000
Cpag 1B9	&	Cpag 38	19.992	24	0.697
Cpag 1C8	&	Cpag 38	10.600	20	0.956
Cpag 2A5B	&	Cpag 38	16.744	24	0.859
Cpag 3A2	&	Cpag 38	12.226	24	0.977
Cpag 3D7	&	Cpag 38	14.491	24	0.935
Cpag 4C1	&	Cpag 38	17.189	24	0.840
Cpag 5D8	&	Cpag 38	12.076	24	0.979
Cpag 6C4B	&	Cpag 38	27.339	24	0.289
Cpag 15	&	Cpag 38	24.488	24	0.434

Appendix 6. F_{ST} values for *Cancer pagurus* in the Irish Sea before (upper diagonal) and after (lower diagonal) ENA correction. Significant F_{ST} values are marked in bold with an asterisk (95% CI).

	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX
AW		0.010	-0.002	-0.001	0.012	**0.015***	0.009	0.011	0.004	0.008	0.006	0.006	0.006
CB	0.010		0.003	-0.006	0.006	-0.001	0.001	0.003	0.004	-0.002	-0.001	-0.003	0.001
LIA	0.003	0.005		-0.003	0.002	0.002	-0.001	-0.001	0.005	0.003	-0.002	0.002	0.003
LIB	0.001	-0.005	0.001		-0.004	-0.005	0.000	-0.002	-0.001	-0.006	-0.006	-0.006	-0.004
LNTZ	0.014	0.006	0.003	-0.002		0.000	0.007	0.004	0.007	0.006	-0.002	0.003	0.005
NW	**0.017***	0.000	0.005	-0.003	0.002		0.004	0.002	**0.010***	0.000	-0.001	0.001	0.002
SW	0.014	0.006	0.002	0.005	**0.012***	0.008		-0.003	0.001	0.004	-0.003	0.002	-0.004
ND	0.011	0.004	0.002	-0.001	0.004	0.002	0.000		0.000	0.000	0.001	-0.002	-0.001
SD	0.005	0.004	**0.008***	0.000	0.008	**0.010***	0.005	0.000		0.000	**0.005***	-0.001	-0.002
WF	0.009	-0.001	0.006	-0.005	0.008	0.001	0.008	0.000	0.001		0.000	-0.003	-0.002
WEX	0.008	0.001	-0.001	-0.004	0.000	0.001	0.000	0.002	**0.006***	0.002		0.001	0.001
M	0.005	-0.002	0.004	-0.005	0.004	0.002	0.006	-0.002	-0.001	-0.003	0.002		0.000
OX	0.006	0.002	**0.006***	-0.003	0.007	0.002	-0.001	-0.001	-0.002	-0.001	0.002	0.000	

Appendix 7. IBD matrix comprising of genetic distance (F_{ST} - upper diagonal) and the corresponding geographic distance (km - lower diagonal) for *Cancer pagurus* in the Irish Sea.

IBD	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX
AW		0.010	-0.002	-0.001	0.012	0.015	0.009	0.011	0.004	0.008	0.006	0.006	0.006
CB	117.327		0.003	-0.006	0.006	-0.001	0.001	0.003	0.004	-0.002	-0.001	-0.003	0.001
LIA	293.576	205.874		-0.003	0.002	0.002	-0.001	-0.001	0.005	0.003	-0.002	0.002	0.003
LIB	293.576	205.874	0.161		-0.004	-0.005	0.000	-0.002	-0.001	-0.006	-0.006	-0.006	-0.004
LNTZ	295.393	204.524	3.923	3.923		0.000	0.007	0.004	0.007	0.006	-0.002	0.003	0.005
NW	28.559	129.880	278.306	278.306	278.570		0.004	0.002	0.010	0.000	-0.001	0.001	0.002
SW	311.334	217.111	57.199	57.199	54.364	301.130		-0.003	0.001	0.004	-0.003	0.002	-0.004
ND	113.490	178.938	282.021	282.021	278.023	141.367	306.590		0.000	-0.001	0.001	-0.002	-0.001
SD	106.906	159.982	251.620	251.620	247.469	134.220	277.193	31.524		0.000	0.005	-0.001	-0.002
WF	215.245	200.043	193.547	193.547	193.893	237.987	213.205	198.584	168.426		0.000	-0.003	-0.002
WEX	165.821	154.400	163.622	163.622	163.081	187.245	182.556	147.342	115.904	48.425		0.001	0.001
M	321.622	230.388	67.510	67.510	64.693	289.152	12.166	308.174	281.867	222.587	188.416		0.000
OX	311.334	217.111	57.199	57.199	54.364	301.130	0.161	306.590	277.193	213.205	182.556	12.121	

Appendix 8. Linkage disequilibrium in *Cancer pagurus* in the NE Atlantic.

Locus pair			Chi^2	df	P-value
Cpag 4	&	Cpag 1B9	42.427	44	0.539
Cpag 4	&	Cpag 2A5B	23.812	44	0.994
Cpag 1B9	&	Cpag 2A5B	61.001	48	0.099
Cpag 4	&	Cpag 3A2	31.554	44	0.920
Cpag 1B9	&	Cpag 3A2	55.436	48	0.215
Cpag 2A5B	&	Cpag 3A2	40.240	48	0.779
Cpag 4	&	Cpag 3D7	26.515	44	0.983
Cpag 1B9	&	Cpag 3D7	55.645	48	0.209
Cpag 2A5B	&	Cpag 3D7	32.961	48	0.952
Cpag 3A2	&	Cpag 3D7	59.706	48	0.120
Cpag 4	&	Cpag 5D8	Infinity	44	Highly sign.
Cpag 1B9	&	Cpag 5D8	Infinity	46	Highly sign.
Cpag 2A5B	&	Cpag 5D8	37.921	46	0.796
Cpag 3A2	&	Cpag 5D8	45.052	46	0.512
Cpag 3D7	&	Cpag 5D8	32.990	46	0.925
Cpag 4	&	Cpag 6C4B	29.999	44	0.947
Cpag 1B9	&	Cpag 6C4B	47.182	48	0.506
Cpag 2A5B	&	Cpag 6C4B	31.856	48	0.965
Cpag 3A2	&	Cpag 6C4B	38.528	48	0.834
Cpag 3D7	&	Cpag 6C4B	46.680	48	0.527
Cpag 5D8	&	Cpag 6C4B	45.302	46	0.501
Cpag 4	&	Cpag 15	42.925	44	0.518
Cpag 1B9	&	Cpag 15	56.498	48	0.187
Cpag 2A5B	&	Cpag 15	52.688	48	0.298
Cpag 3A2	&	Cpag 15	54.013	48	0.256
Cpag 3D7	&	Cpag 15	63.303	48	0.068
Cpag 5D8	&	Cpag 15	58.919	46	0.096
Cpag 6C4B	&	Cpag 15	51.901	48	0.324

Appendix 9. F_{ST} values for *Cancer pagurus* in the NE Atlantic before (upper diagonal) and after (lower diagonal) ENA correction. Significant F_{ST} values are marked in bold with an asterisk (95% CI).

	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX	NNW	NSE	NSW	NAB	Pen	7Brt	6Jer	6Has	6Har	6Sea	NNO
AW		0.006	-0.003	-0.003	0.017	0.011	0.004	0.010	0.002	0.003	0.006	0.005	0.003	**0.014***	0.005	**0.028***	0.004	-0.001	-0.001	0.007	0.003	0.006	0.004	0.016
CB	0.007		0.003	-0.004	0.006	0.000	0.000	0.004	0.006	-0.001	-0.001	-0.002	0.001	**0.010***	0.004	0.008	0.009	0.006	0.003	0.006	0.005	0.004	0.004	0.006
LIA	0.000	0.003		-0.001	0.004	0.001	-0.002	-0.002	0.005	0.002	-0.001	0.002	0.003	**0.014***	0.008	**0.023***	**0.010***	0.000	**0.006***	**0.006***	0.005	0.005	0.004	0.011
LIB	-0.001	-0.003	0.001		-0.004	-0.005	0.001	-0.001	0.001	-0.006	-0.005	-0.005	-0.003	0.002	0.001	0.004	-0.008	-0.002	-0.003	-0.002	-0.003	-0.005	-0.005	0.000
LNTZ	0.018	0.006	0.003	-0.001		-0.001	0.008	0.002	0.008	0.006	0.000	0.003	0.005	**0.019***	0.023	**0.016***	**0.013***	0.006	0.012	0.006	0.008	0.007	0.002	0.009
NW	0.013	0.000	0.001	-0.003	0.000		0.004	0.003	**0.013***	0.002	-0.001	0.002	0.003	0.007	0.004	**0.012***	**0.008***	**0.008***	0.005	**0.008***	**0.008***	**0.007***	**0.007***	**0.011***
SW	0.007	0.003	0.000	0.005	0.010	0.006		-0.004	-0.001	0.002	-0.002	-0.001	-0.006	0.002	0.003	0.010	0.011	0.000	0.001	0.001	-0.002	0.000	-0.001	0.003
ND	0.010	0.005	-0.001	0.000	0.001	0.002	-0.003		0.001	0.000	0.001	-0.002	-0.001	0.008	0.007	0.016	**0.008***	0.000	0.003	0.001	0.000	0.000	-0.001	0.000
SD	0.002	0.006	0.006	0.001	0.008	**0.013***	0.001	0.001		0.001	**0.005***	-0.002	-0.001	0.010	0.016	**0.018***	0.007	-0.001	0.000	0.003	-0.002	0.001	-0.001	0.002
WF	0.003	-0.001	0.003	-0.005	0.006	0.002	0.004	0.000	0.001		-0.001	-0.003	-0.002	0.001	0.005	0.008	0.003	0.000	0.000	0.001	0.000	0.000	0.001	0.001
WEX	0.007	0.000	0.000	-0.004	0.000	-0.001	-0.001	0.001	**0.006***	0.000		0.001	0.001	0.006	0.009	0.008	0.008	0.003	0.003	0.003	0.003	0.003	0.005	0.007
M	0.005	-0.002	0.002	-0.004	0.004	0.002	0.000	-0.003	-0.001	-0.003	0.001		-0.001	0.008	0.007	0.013	0.005	-0.001	0.000	0.002	-0.001	-0.001	-0.002	-0.002
OX	0.003	0.001	0.004	-0.002	0.006	0.003	-0.005	-0.001	-0.001	-0.002	0.001	-0.001		0.002	0.005	0.008	0.008	0.001	-0.001	0.000	-0.001	0.000	-0.001	0.002
NNW	**0.016***	**0.009***	**0.015***	0.004	**0.017***	0.007	0.003	0.008	**0.011***	0.002	0.006	0.007	0.001		0.003	0.003	**0.010***	**0.013***	0.002	**0.008***	**0.006***	0.005	**0.011***	**0.010***
NSE	0.006	0.004	0.009	0.002	0.021	0.004	0.006	0.007	0.015	0.005	0.009	0.006	0.006	0.004		**0.019***	0.009	0.010	0.002	0.012	0.010	0.008	0.011	0.012
NSW	**0.028***	0.007	**0.022***	0.004	**0.016***	**0.010***	0.011	0.015	**0.017***	0.008	0.008	0.011	0.007	0.003	**0.018***		**0.019***	**0.024***	**0.014***	0.009	**0.013***	**0.011***	**0.016***	0.014
NAB	0.006	0.009	**0.011***	-0.006	**0.013***	**0.009***	**0.013***	**0.008***	0.007	0.004	0.008	0.005	0.008	**0.011***	0.009	**0.018***		0.005	0.003	0.009	0.003	0.003	0.002	0.005
Pen	0.001	0.006	0.001	-0.001	0.005	**0.008***	0.002	0.000	-0.001	0.000	0.003	-0.001	0.002	**0.014***	0.010	**0.023***	0.006		0.002	0.004	0.001	0.001	-0.002	0.002
7Brt	0.000	0.003	**0.005***	-0.003	0.011	0.005	0.002	0.003	0.000	0.000	0.003	0.000	-0.001	0.003	0.002	**0.012***	0.003	0.002		0.003	0.000	0.002	0.002	0.003
6Jer	0.007	**0.007***	**0.007***	-0.001	0.007	**0.009***	0.003	0.001	0.003	0.001	0.004	0.002	0.000	**0.008***	0.012	0.008	0.008	0.004	0.003		0.001	0.003	0.001	0.006
6Has	0.003	0.005	**0.006***	-0.002	0.007	**0.008***	-0.001	0.000	-0.002	0.000	0.004	-0.001	-0.001	**0.006***	0.010	**0.011***	0.003	0.001	0.000	0.001		0.000	-0.002	0.000
6Har	0.007	**0.004***	0.006	-0.004	0.007	**0.007***	0.001	0.001	0.001	0.000	0.004	-0.001	0.000	0.005	0.008	**0.010***	0.003	0.002	0.002	0.003	0.000		-0.002	0.000
6Sea	0.005	0.005	0.005	-0.004	0.003	**0.008***	0.001	-0.001	-0.001	0.001	0.005	-0.002	-0.001	**0.011***	0.010	**0.015***	0.002	-0.001	0.003	0.002	-0.002	-0.001		-0.002
NNO	0.015	0.006	0.012	0.000	0.009	**0.011***	0.005	0.000	0.002	0.001	0.008	-0.002	0.002	**0.010***	0.011	0.013	0.005	0.002	0.003	0.006	0.000	0.000	-0.002	

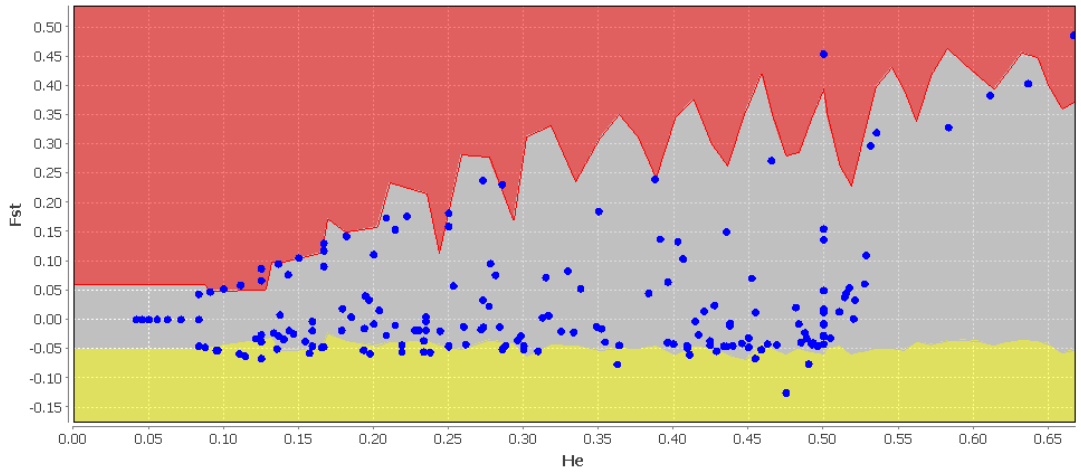
Appendix 10. IBD matrix comprising of genetic distance (F_{ST} - upper diagonal) and the corresponding geographic distance (km - lower diagonal) for *Cancer pagurus* in the NE Atlantic.

IBD	AW	CB	LIA	LIB	LNTZ	NW	SW	ND	SD	WF	WEX	M	OX	NNW	NSE	NSW	NAB	Pen	7Brt	6Jer	6Has	6Har	6Sea	NNO
AW		0.006	-0.003	-0.003	0.017	0.011	0.004	0.010	0.002	0.003	0.006	0.005	0.003	0.014	0.005	0.028	0.004	-0.001	-0.001	0.007	0.003	0.006	0.004	0.016
CB	117.327		0.003	-0.004	0.006	0.000	0.000	0.004	0.006	-0.001	-0.001	-0.002	0.001	0.010	0.004	0.008	0.009	0.006	0.003	0.006	0.005	0.004	0.004	0.006
LIA	293.576	205.874		-0.001	0.004	0.001	-0.002	-0.002	0.005	0.002	-0.001	0.002	0.003	0.014	0.008	0.023	0.010	0.000	0.006	0.006	0.005	0.005	0.004	0.011
LIB	293.576	205.874	0.161		-0.004	-0.005	0.001	-0.001	0.001	-0.006	-0.005	-0.005	-0.003	0.002	0.001	0.004	-0.008	-0.002	-0.003	-0.002	-0.003	-0.005	-0.005	0.000
LNTZ	295.393	204.524	3.923	3.923		-0.001	0.008	0.002	0.008	0.006	-0.001	0.003	0.005	0.019	0.023	0.016	0.013	0.006	0.012	0.006	0.008	0.007	0.002	0.009
NW	28.559	129.880	278.306	278.306	278.570		0.004	0.003	0.013	0.002	-0.001	0.002	0.003	0.007	0.004	0.012	0.008	0.008	0.005	0.008	0.008	0.007	0.007	0.011
SW	311.334	217.111	57.199	57.199	54.364	301.130		-0.004	-0.001	0.002	-0.002	-0.001	-0.006	0.002	0.003	0.010	0.011	0.000	0.001	0.001	-0.002	0.000	-0.001	0.003
ND	113.490	178.938	282.021	282.021	278.023	141.367	306.590		0.001	0.000	0.001	-0.002	-0.001	0.008	0.007	0.016	0.008	0.000	0.003	0.001	0.000	0.000	-0.001	0.000
SD	106.906	159.982	251.620	251.620	247.469	134.220	277.193	31.524		0.001	0.005	-0.002	-0.001	0.010	0.016	0.018	0.007	-0.001	0.000	0.003	-0.002	0.001	-0.001	0.002
WF	215.245	200.043	193.547	193.547	193.893	237.987	213.205	198.584	168.426		-0.001	-0.003	-0.002	0.001	0.005	0.008	0.003	0.000	0.000	0.001	0.000	0.000	0.001	0.001
WEX	165.821	154.400	163.622	163.622	163.081	187.245	182.556	147.342	115.904	48.425		0.001	0.001	0.006	0.009	0.008	0.008	0.003	0.003	0.003	0.003	0.003	0.005	0.007
M	321.622	230.388	67.510	67.510	64.693	289.152	12.166	308.174	281.867	222.587	188.416		-0.001	0.008	0.007	0.013	0.005	-0.001	0.000	0.002	-0.001	-0.001	-0.002	-0.002
OX	311.334	217.111	57.199	57.199	54.364	301.130	0.161	306.590	277.193	213.205	182.556	12.121		0.002	0.005	0.008	0.008	0.001	-0.001	0.000	-0.001	0.000	-0.001	0.002
NNW	522.629	617.795	761.694	761.694	757.632	559.159	766.010	492.215	515.240	674.555	622.996	767.377	766.010		0.003	0.003	0.010	0.013	0.002	0.008	0.006	0.005	0.011	0.010
NSE	215.245	200.043	193.547	193.547	193.893	237.987	213.205	198.584	168.426	0.161	48.425	222.587	213.205	674.555		0.019	0.009	0.010	0.002	0.012	0.010	0.008	0.011	0.012
NSW	754.048	675.595	619.169	619.169	625.134	707.949	654.929	715.483	742.235	483.850	519.700	666.070	654.929	319.154	483.850		0.019	0.024	0.014	0.009	0.013	0.011	0.016	0.014
NAB	117.327	0.161	205.874	205.874	204.524	129.880	217.111	178.938	159.982	200.043	154.400	230.388	217.111	617.795	200.043	675.595		0.005	0.003	0.009	0.003	0.003	0.002	0.005
Pen	351.681	298.813	134.437	134.437	138.701	368.223	191.297	371.483	340.197	235.588	230.315	201.745	191.297	790.439	235.588	588.436	298.813		0.002	0.004	0.001	0.001	-0.002	0.002
7Brt	556.769	501.426	339.412	339.412	343.211	572.229	394.137	573.761	541.647	438.388	432.494	403.587	394.137	984.577	438.388	781.048	501.426	201.409		0.003	0.000	0.002	0.002	0.003
6Jer	639.466	586.949	422.337	422.337	428.523	656.683	479.087	659.039	625.686	522.271	515.564	487.011	479.087	1074.928	522.271	867.609	586.949	285.789	146.103		0.001	0.003	0.001	0.006
6Has	829.066	777.188	612.217	612.217	617.336	845.624	669.527	848.394	814.891	710.930	704.382	674.511	669.527	1262.465	710.930	1059.214	777.188	474.662	398.267	269.702		0.000	-0.002	0.000
6Har	988.066	936.030	769.267	769.267	777.379	1005.909	827.053	1005.979	974.029	868.833	862.152	832.750	827.053	1420.733	868.833	1217.477	936.030	631.804	556.244	425.830	158.303		-0.002	0.000
6Sea	1220.939	1308.462	1229.070	1229.070	1237.128	1229.504	1286.504	1156.775	1186.349	1328.458	1289.044	1292.119	1286.504	1045.651	1328.458	1266.132	1308.462	1091.783	1014.452	885.590	619.107	467.741		-0.002
NNO	1467.259	1534.512	1661.506	1661.506	1673.436	1456.692	1655.117	1388.475	1419.124	1574.298	1520.243	1669.160	1655.117	1282.465	1574.298	1497.143	1534.512	1743.069	1771.871	1642.140	1373.763	1218.158	889.476	

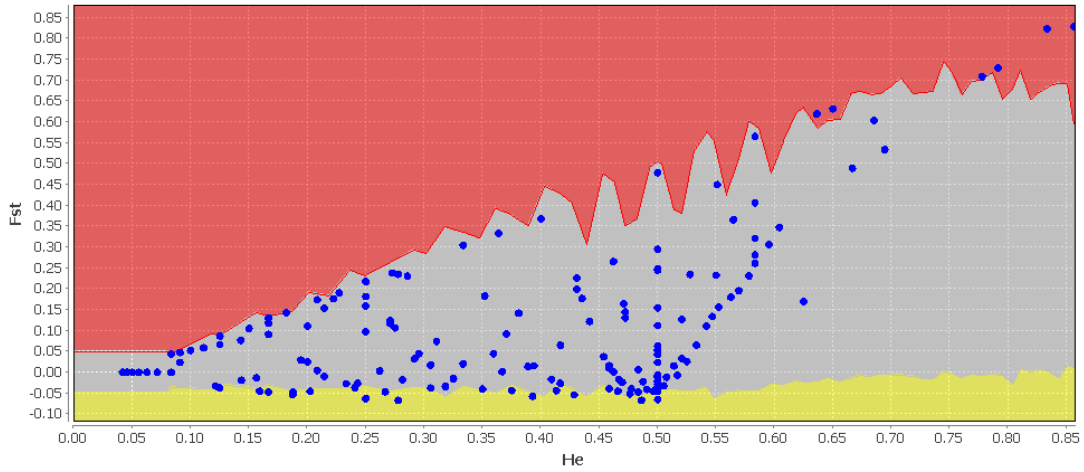
Appendix 11. LOSITAN divergent loci list with heterozygosity (Het), F_{ST} and P value (simulated F_{ST} < sample F_{ST}) of outlier test (P > 0.995). Loci status after false discovery rate (FDR) correction and corresponding BayeScan loci are shown.

Locus	Het	F_{ST}	P	FDR	BayeScan
RADid_0075953_depth_98.0000000074	0.504	0.253	1.000*	Outlier	541*
RADid_0020714_depth_95.0000000016	0.329	0.227	1.000*	Outlier	148*
RADid_0060888_depth_210.0000000052	0.496	0.220	1.000*	Outlier	469*
RADid_0039681_depth_45.0000000042	0.486	0.202	1.000*	Outlier	358
RADid_0001042_depth_52.0000000031	0.455	0.192	0.999*	Outlier	15
RADid_0052316_depth_35.0000000018	0.446	0.189	0.999*	Outlier	427
RADid_0071628_depth_22.0000000018	0.449	0.178	0.999*	Outlier	513
RADid_0029759_depth_109.0000000024	0.412	0.173	0.998*	Outlier	263
RADid_0009374_depth_40.0000000052	0.508	0.171	0.997*	-	82
RADid_0077263_depth_118.0000000056	0.507	0.169	0.997*	-	551
RADid_0023491_depth_91.0000000025	0.401	0.159	0.996*	-	190

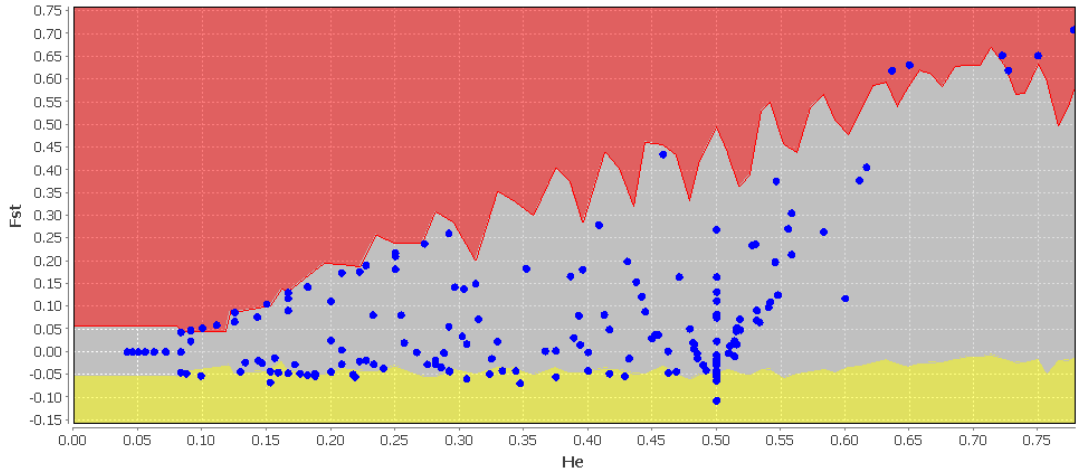
Appendix 12. LOSITAN outlier plots for pairwise comparisons of *Cancer pagurus*.



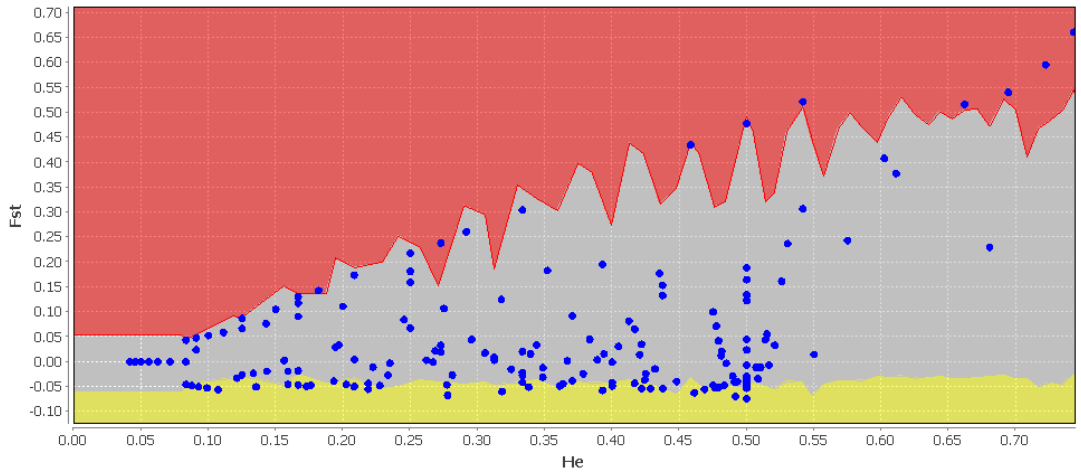
NWI vs. SWI outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 7 SNPs were detected as positive outliers (strong outliers).



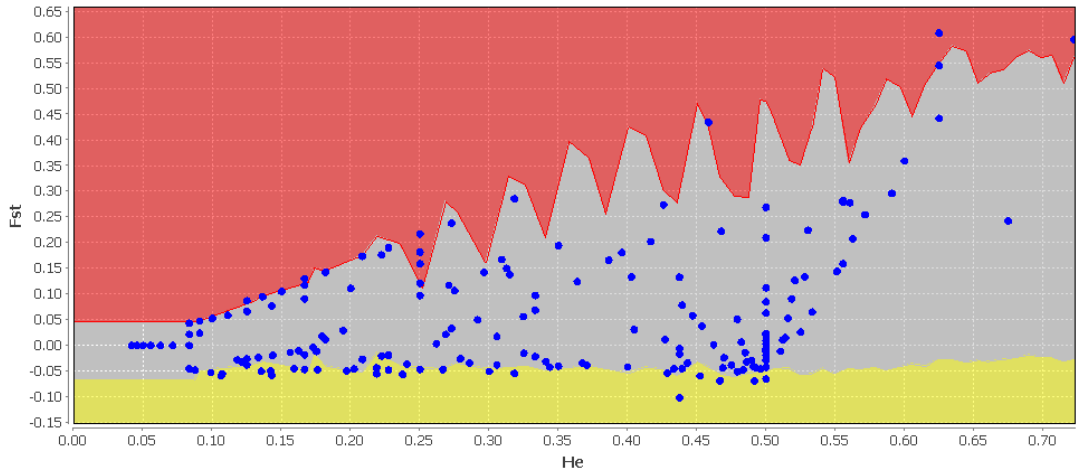
NWI vs. ISA outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 7 SNPs were detected as positive outliers, with 6 SNPs remaining significant outliers after FDR correction (strong outliers).



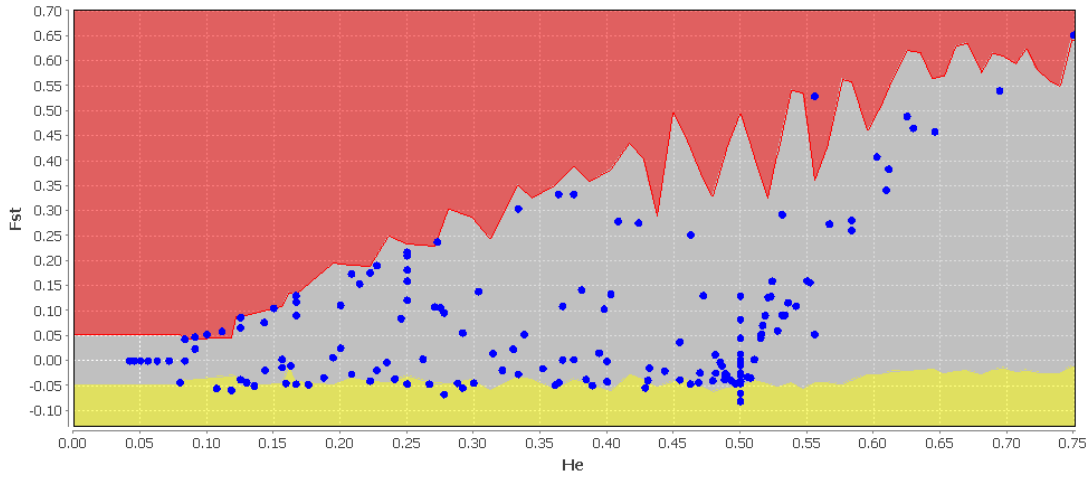
NWI vs. ISB outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 21 SNPs were detected as positive outliers (strong outliers).



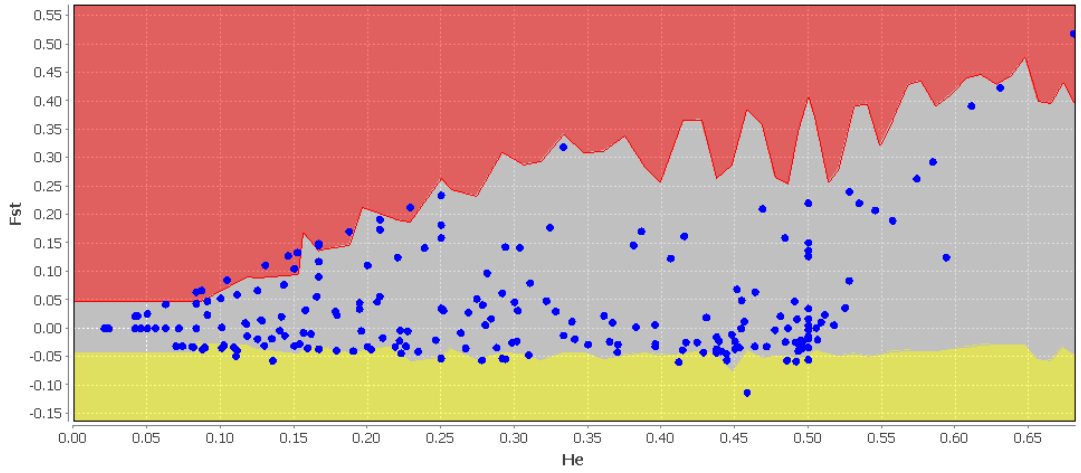
NWI vs. NQ outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 12 SNPs were detected as positive outliers (strong outliers).



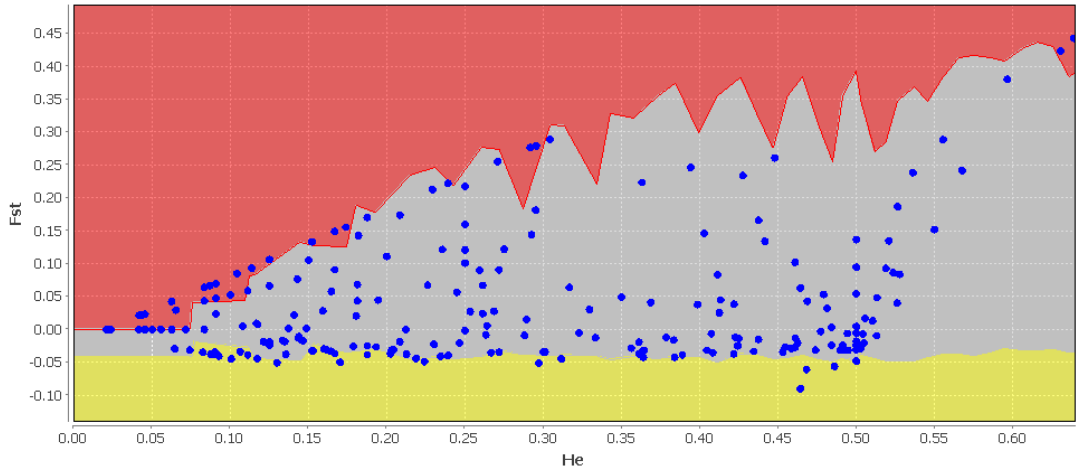
NWI vs. Brit outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 19 SNPs were detected as positive outliers, with 18 SNPs remaining significant outliers after FDR correction (strong outliers).



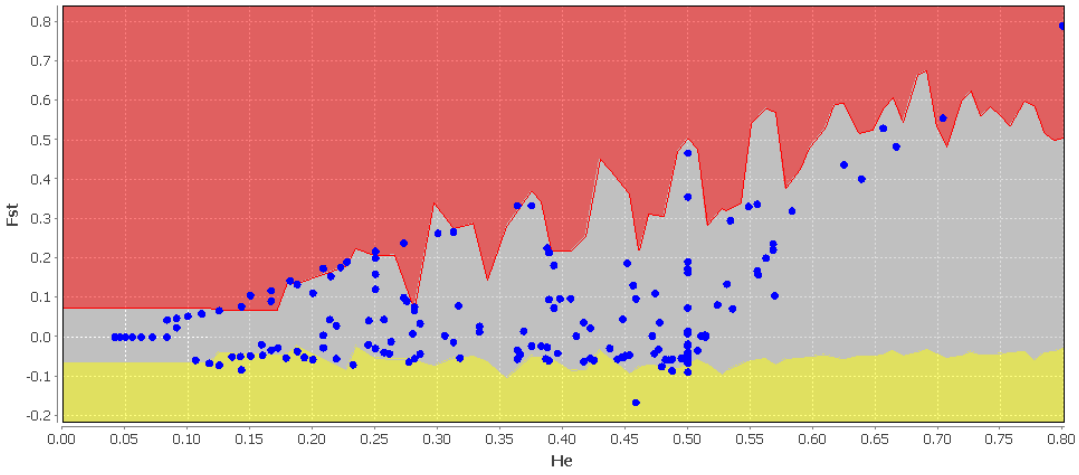
NWI vs. Hast outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 32 SNPs were detected as positive outliers (strong outliers).



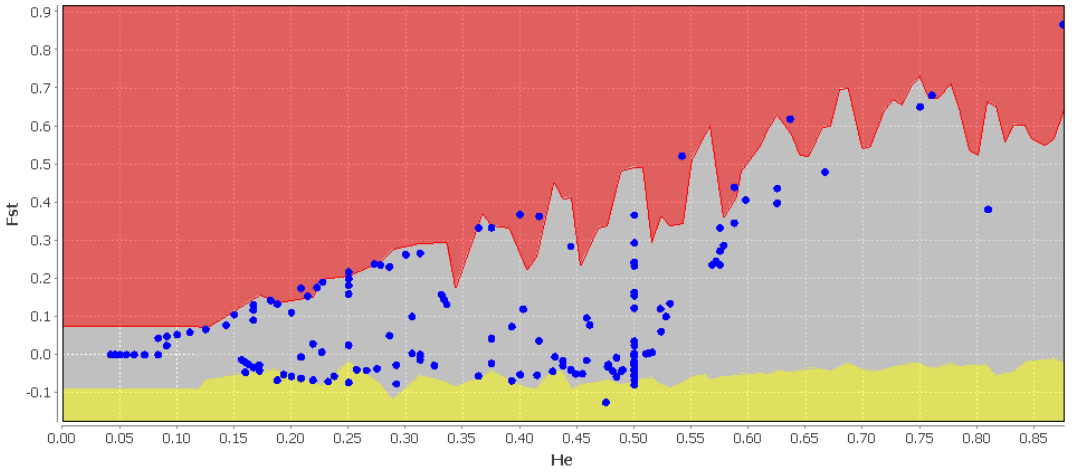
NWI vs. Har outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 10 SNPs were detected as positive outliers (strong outliers).



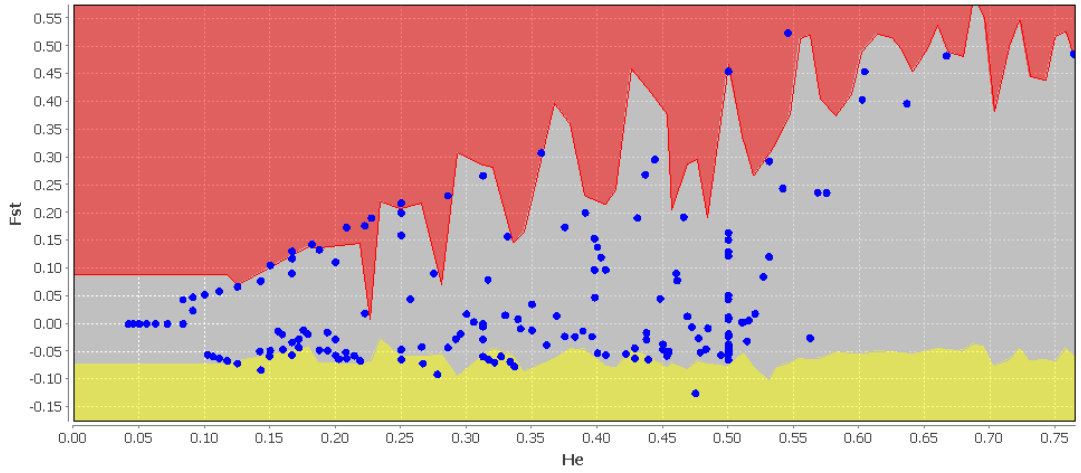
NWI vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 55 SNPs were detected as positive outliers (strong outliers).



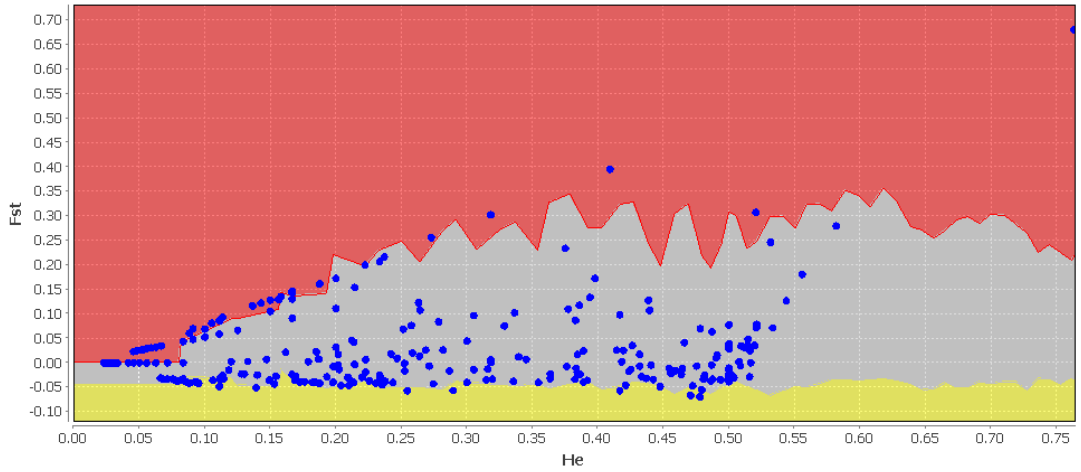
NWI vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 28 SNPs were detected as positive outliers (strong outliers).



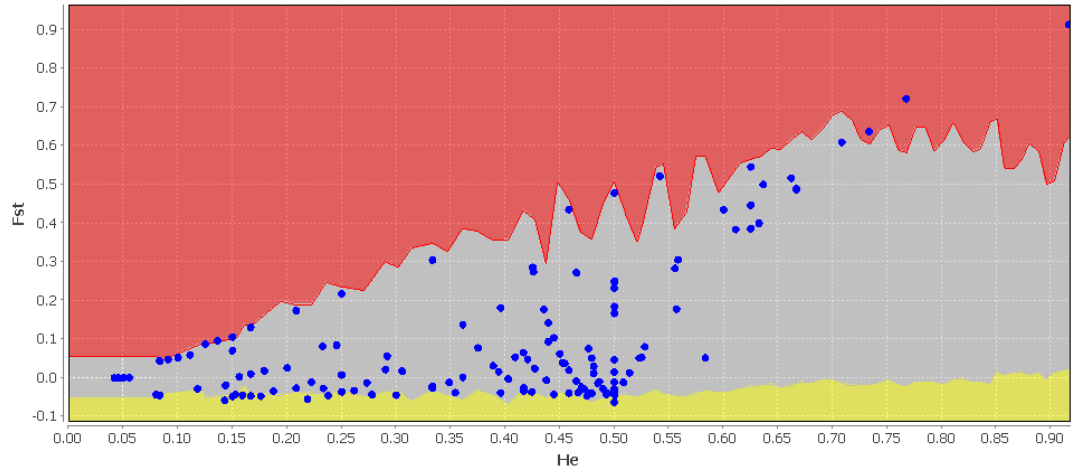
NWI vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 27 SNPs were detected as positive outliers (strong outliers).



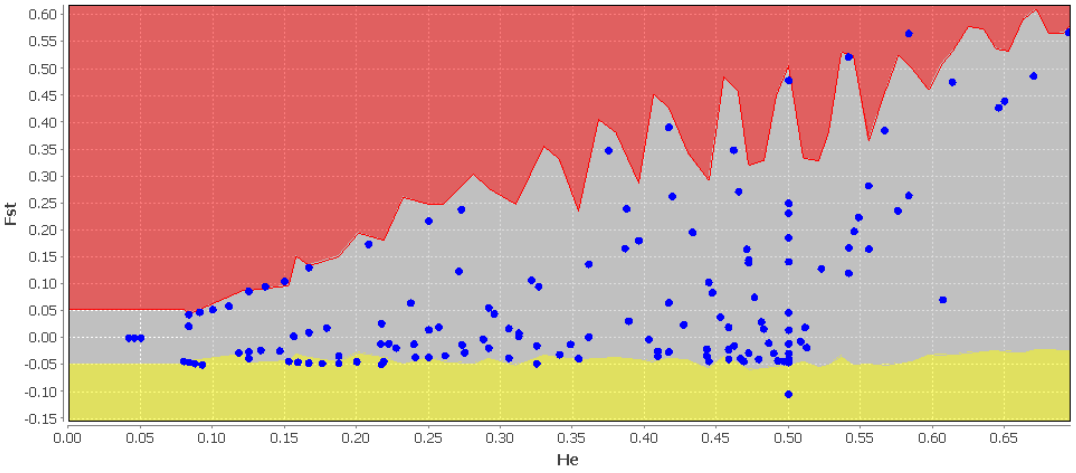
NWI vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 8 SNPs were detected as positive outliers (strong outliers).



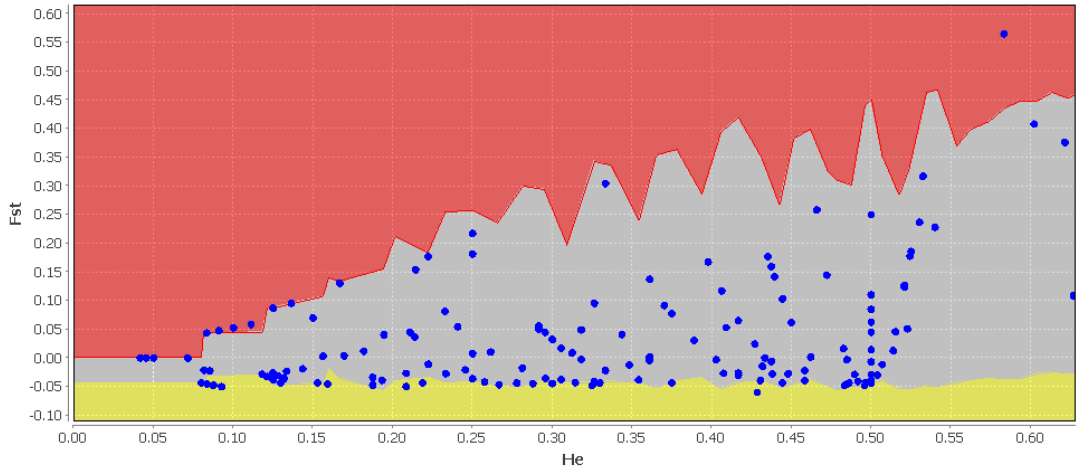
NWI vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 41 SNPs were detected as positive outliers (strong outliers).



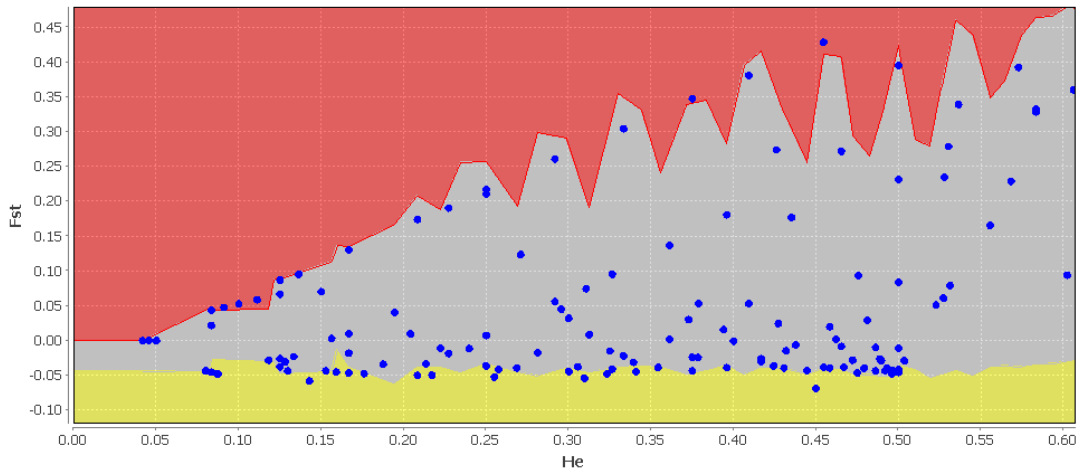
SWI vs. ISA outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 5 SNPs were detected as positive outliers, with 4 SNPs remaining significant outliers after FDR correction (strong outliers).



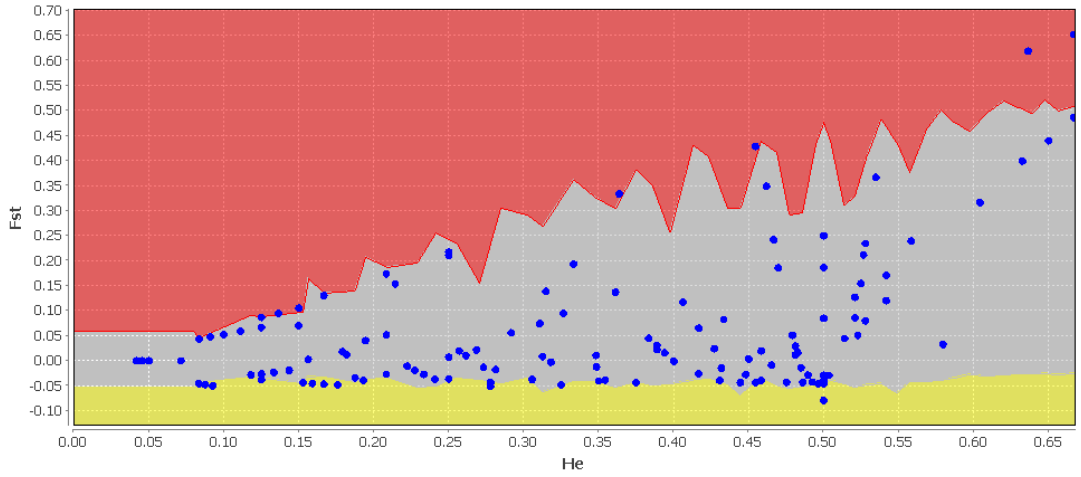
SWI vs. ISB outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 13 SNPs were detected as positive outliers (strong outliers).



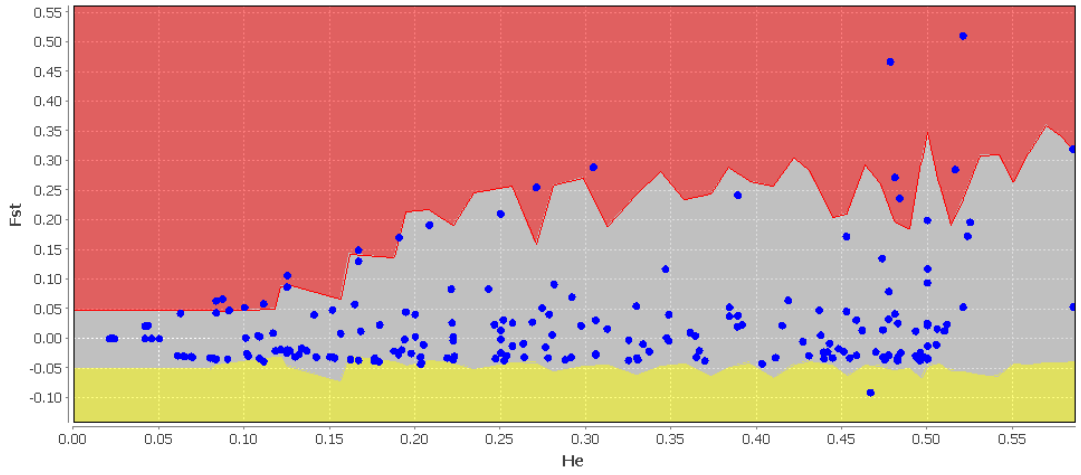
SWI vs. NQ outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 15 SNPs were detected as positive outliers (strong outliers).



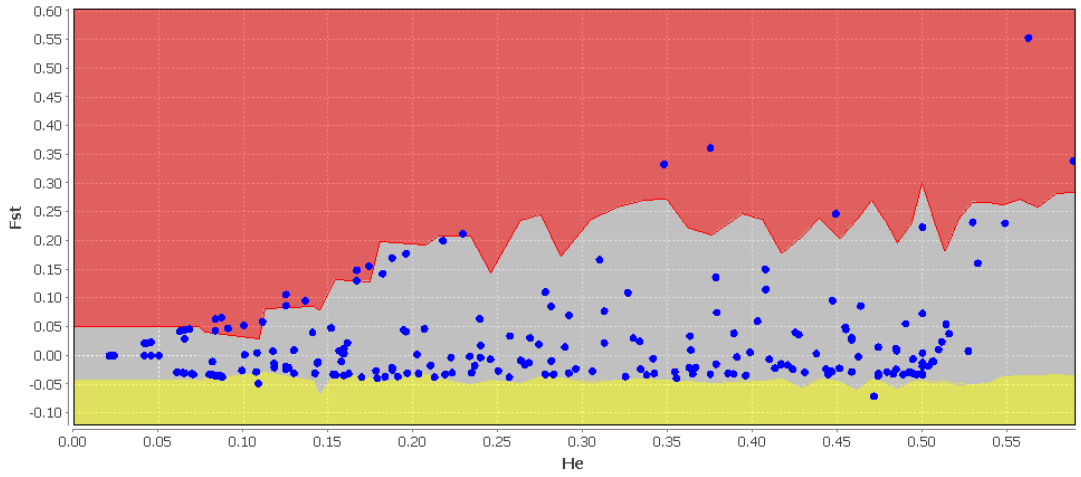
SWI vs. Brit outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 20 SNPs were detected as positive outliers (strong outliers).



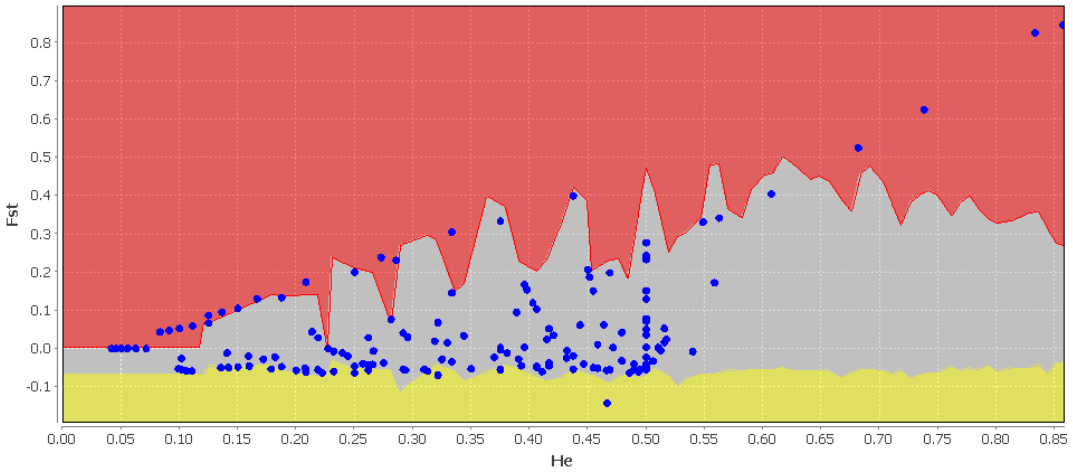
SWI vs. Hast outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 5 SNPs were detected as positive outliers (strong outliers).



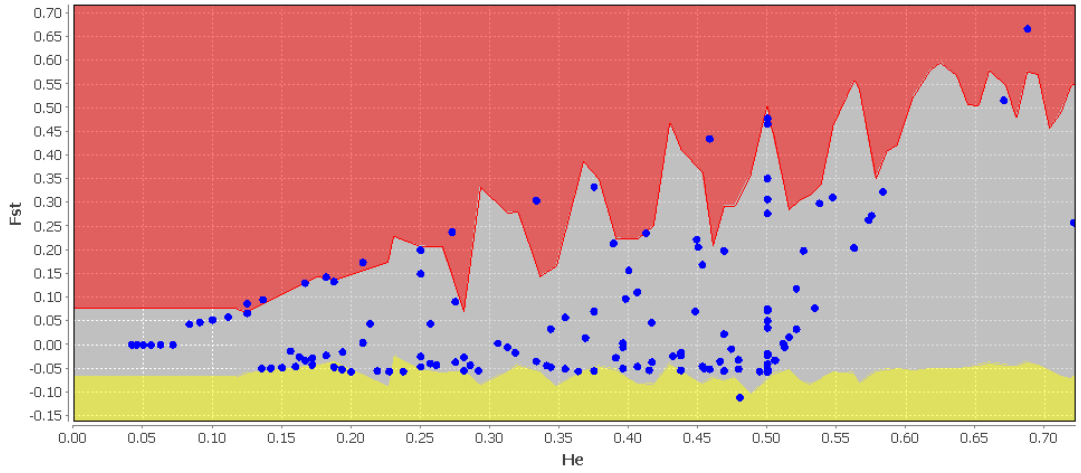
SWI vs. Har outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 16 SNPs were detected as positive outliers (strong outliers).



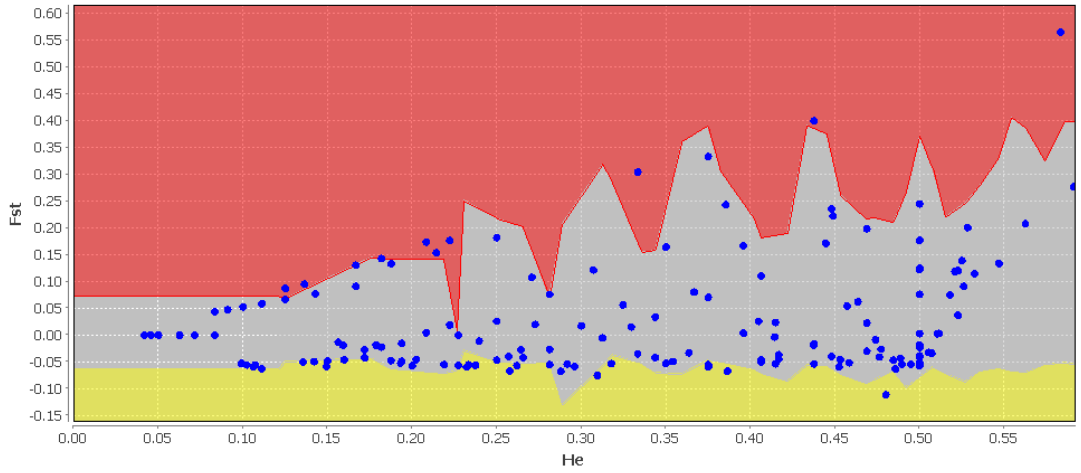
SWI vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 15 SNPs were detected as positive outliers (strong outliers).



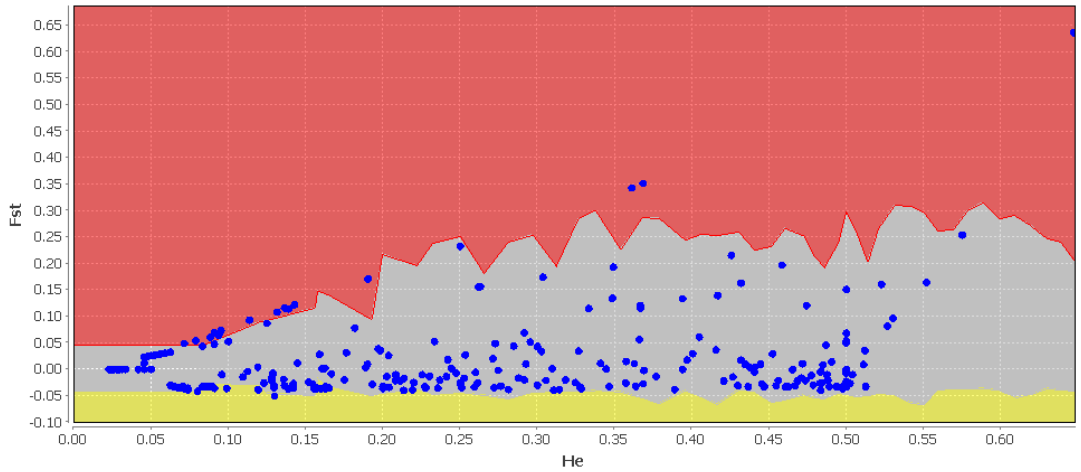
SWI vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 39 SNPs were detected as positive outliers (strong outliers).



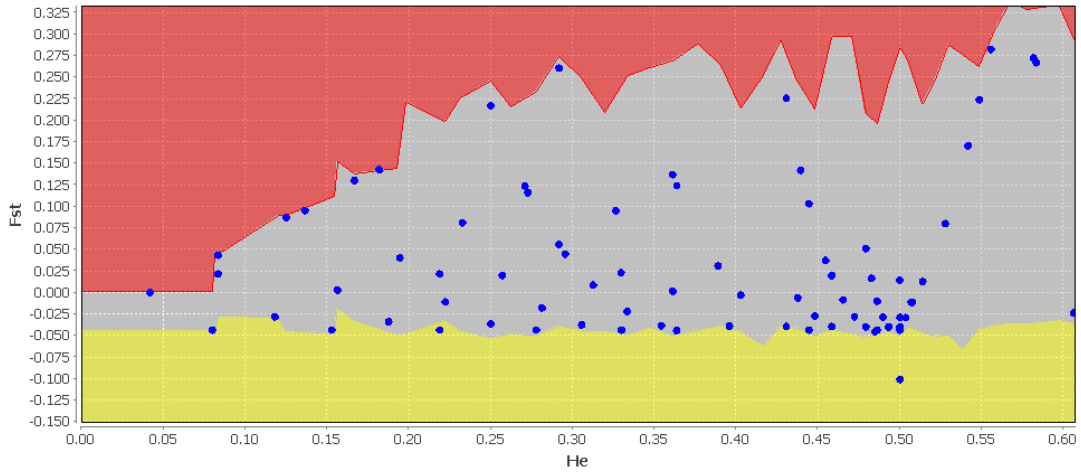
SWI vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 9 SNPs were detected as positive outliers, with 5 SNPs remaining significant outliers after FDR correction (strong outliers).



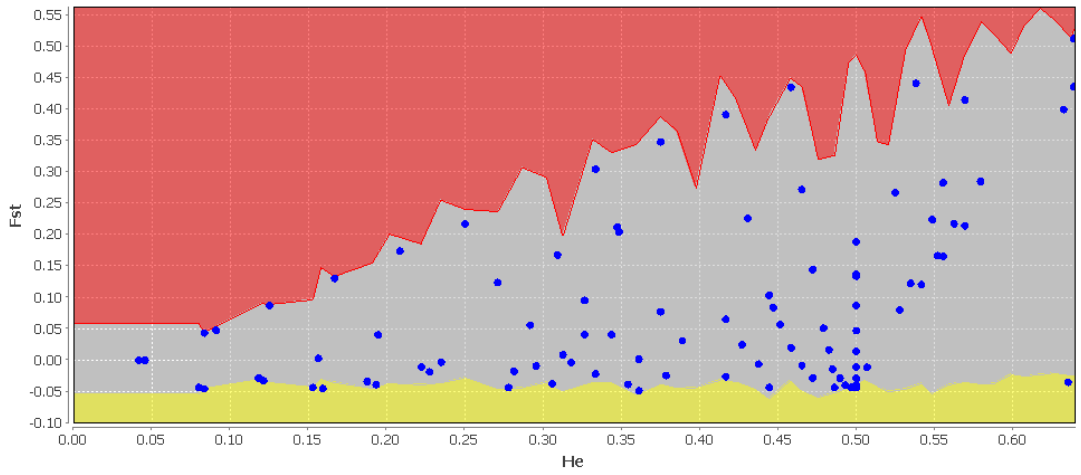
SWI vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 26 SNPs were detected as positive outliers (strong outliers).



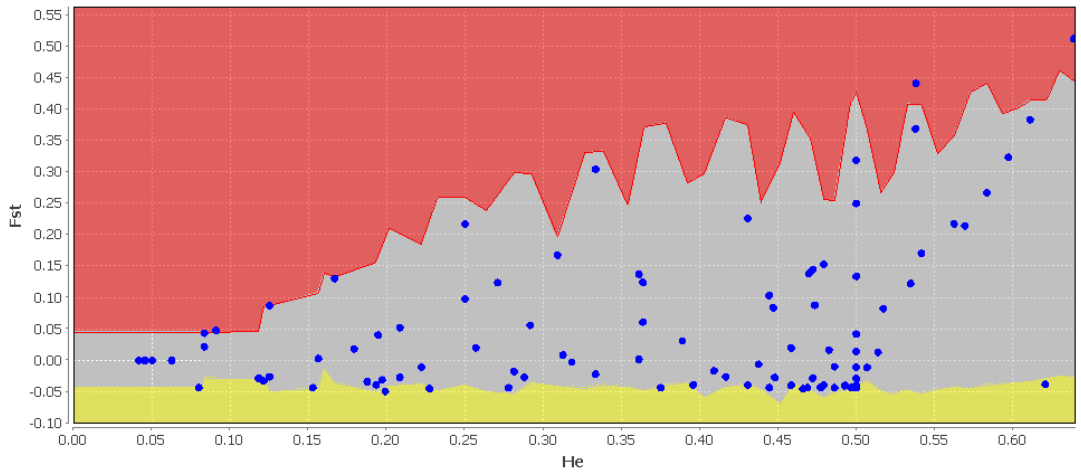
SWI vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 22 SNPs were detected as positive outliers (strong outliers).



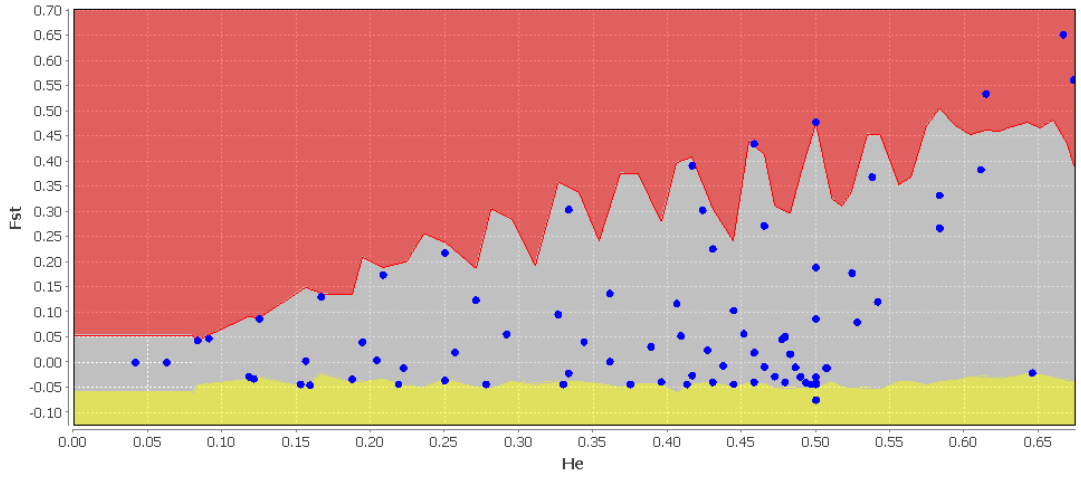
ISA vs. ISB outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 19 SNPs were detected as positive outliers (strong outliers).



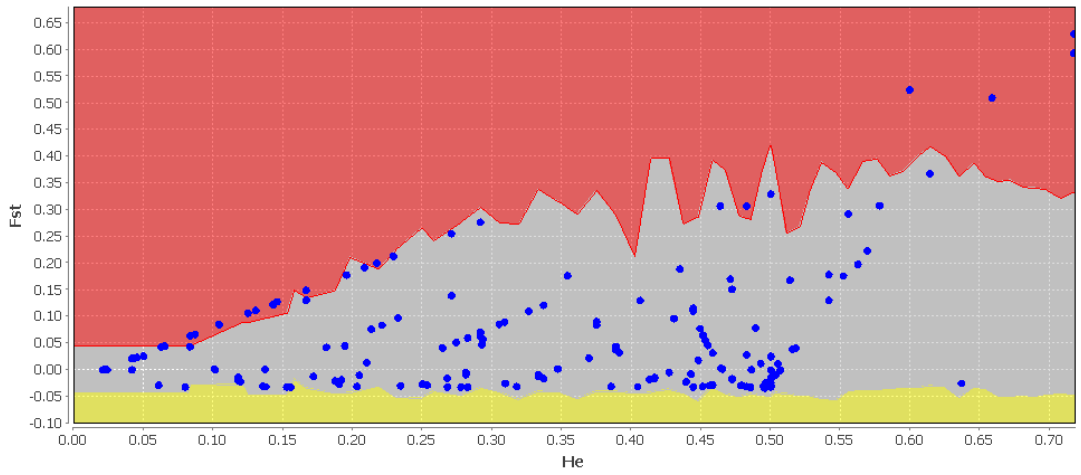
ISA vs. NQ outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). No SNPs were detected as positive outliers.



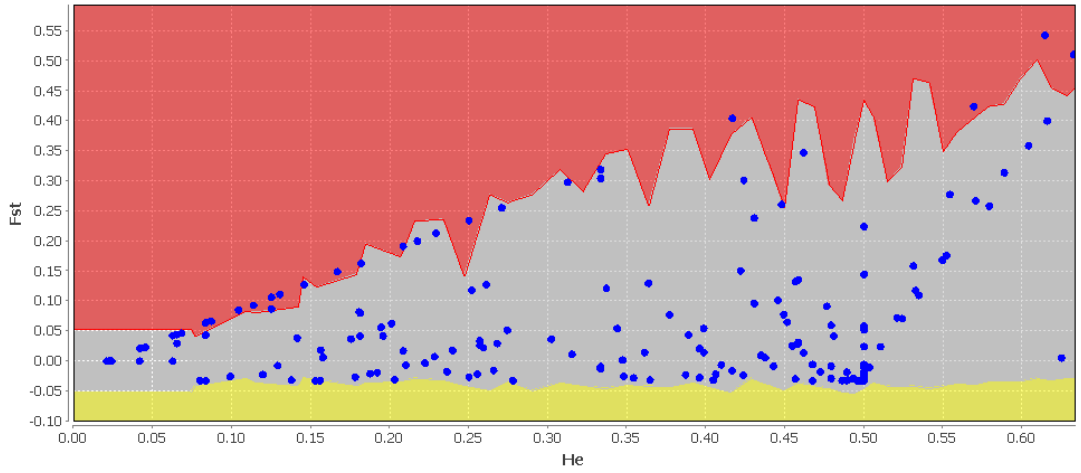
ISA vs. Brit outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 11 SNPs were detected as positive outliers, with 9 SNPs remaining significant outliers after FDR correction (strong outliers).



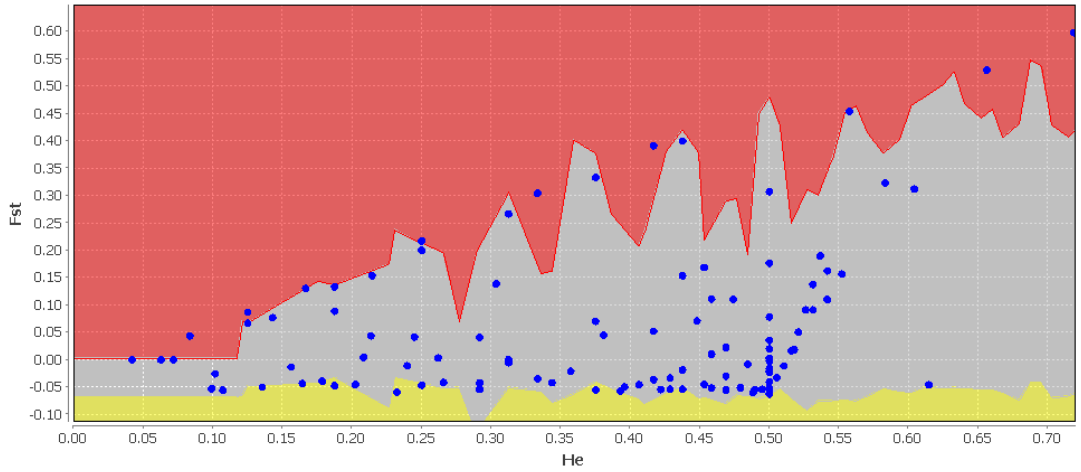
ISA vs. Hast outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 5 SNPs were detected as positive outliers (strong outliers).



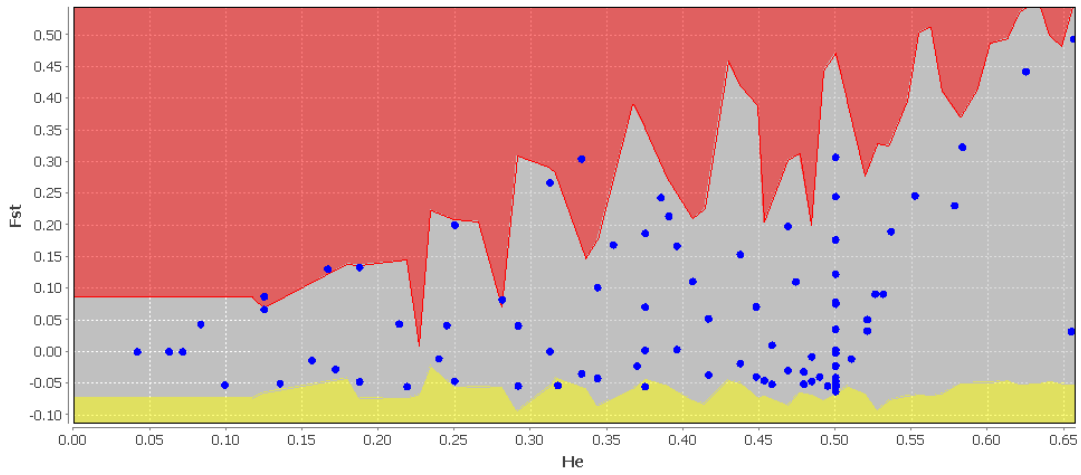
ISA vs. Har outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 18 SNPs were detected as positive outliers, with 15 SNPs remaining significant outliers after FDR correction (strong outliers).



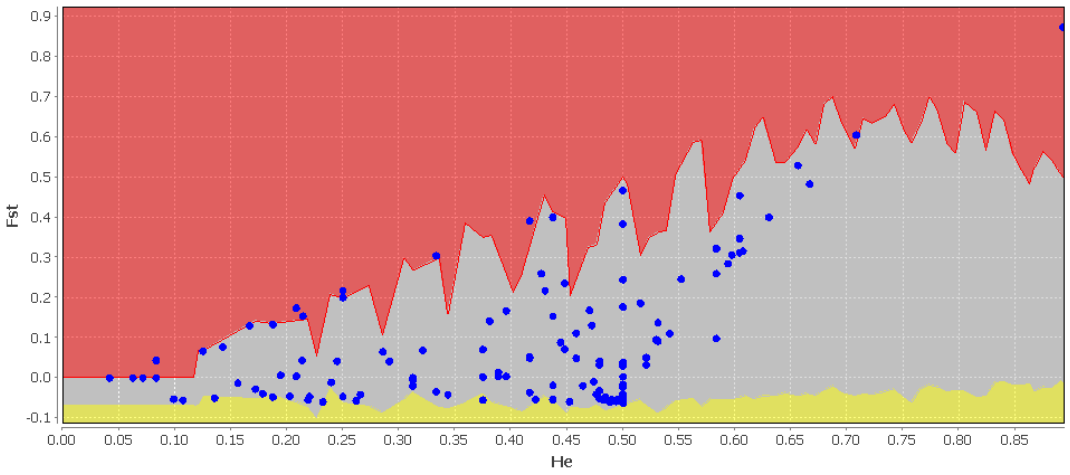
ISA vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 16 SNPs were detected as positive outliers, with 13 SNPs remaining significant outliers after FDR correction (strong outliers).



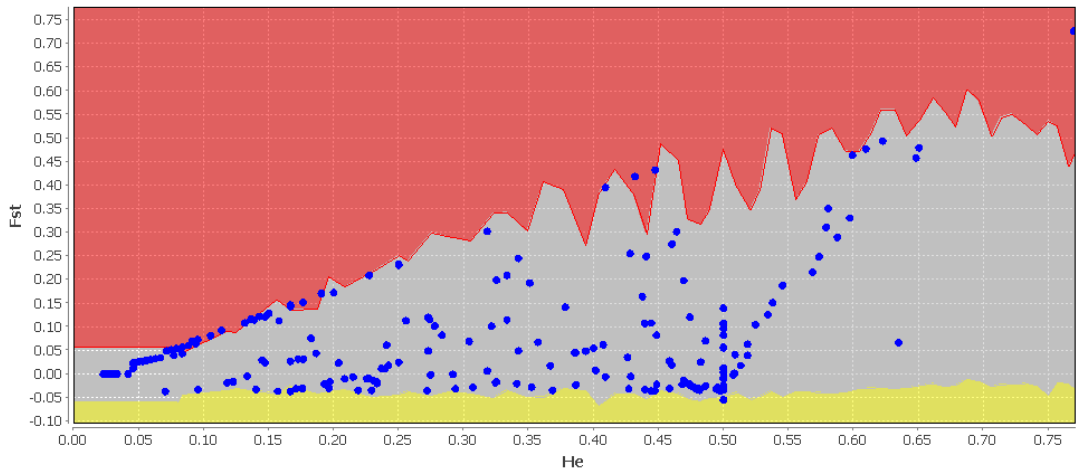
ISA vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 15 SNPs were detected as positive outliers (strong outliers).



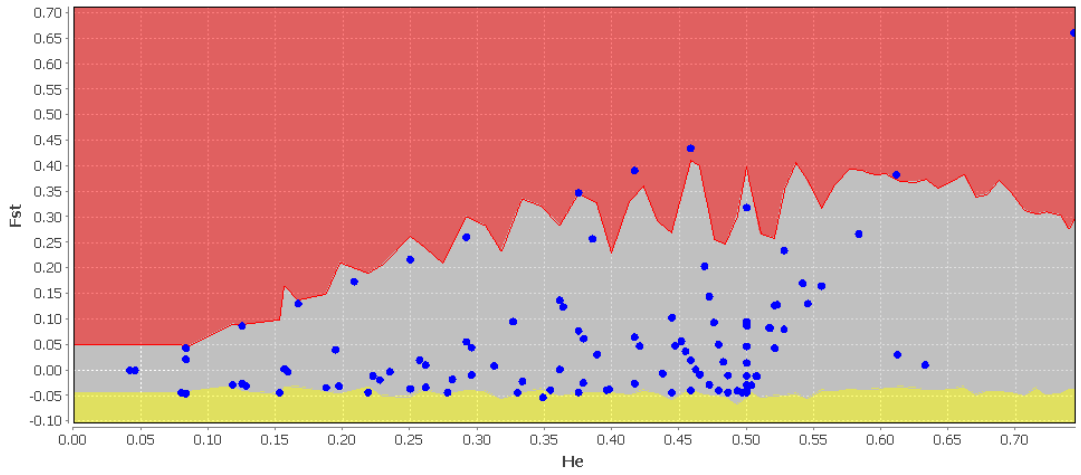
ISA vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 3 SNPs were detected as positive outliers (strong outliers).



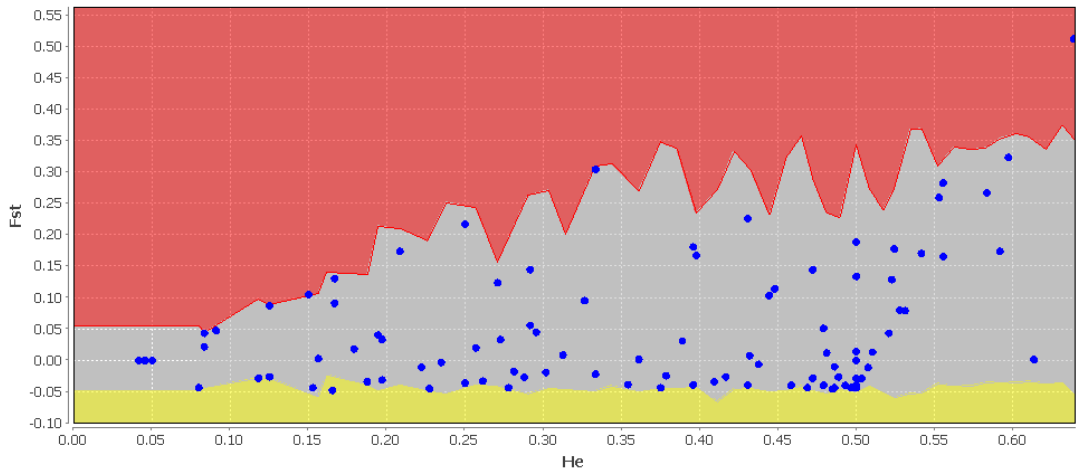
ISA vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 13 SNPs were detected as positive outliers (strong outliers).



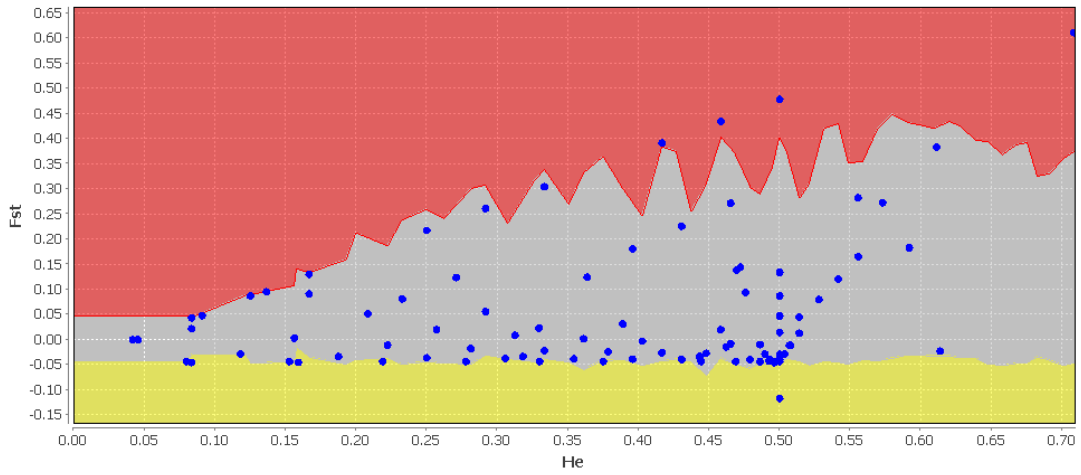
ISA vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 15 SNPs were detected as positive outliers, with 14 SNPs remaining significant outliers after FDR correction (strong outliers).



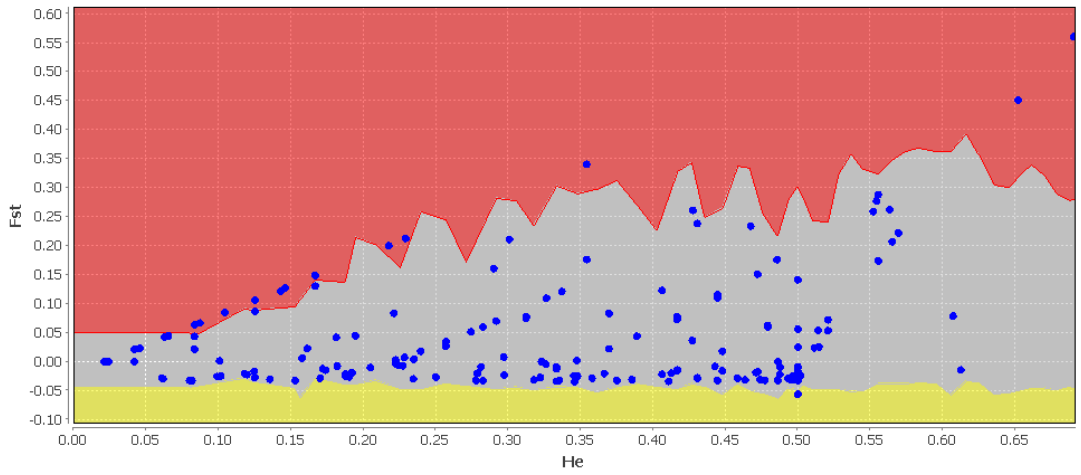
ISB vs. NQ outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 5 SNPs were detected as positive outliers (strong outliers).



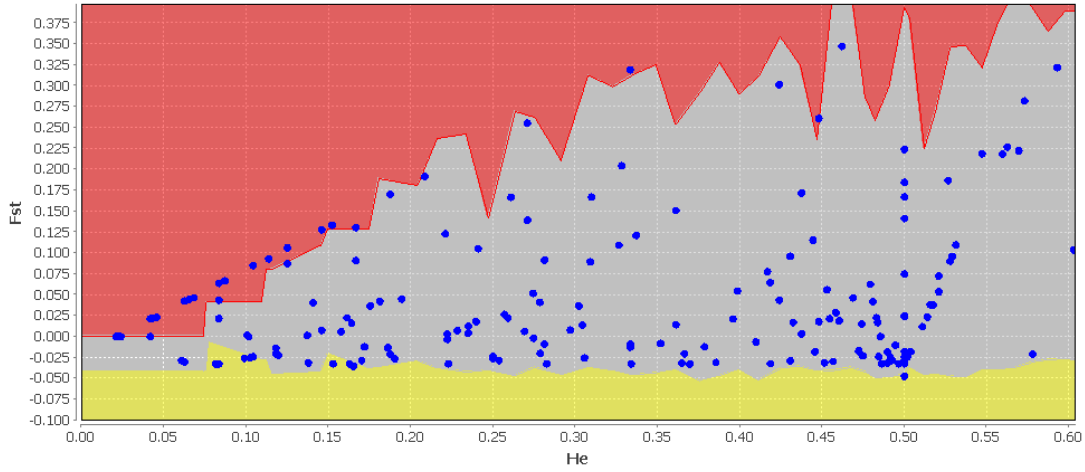
ISB vs. Brit outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 1 SNP was detected as a positive outlier (strong outlier).



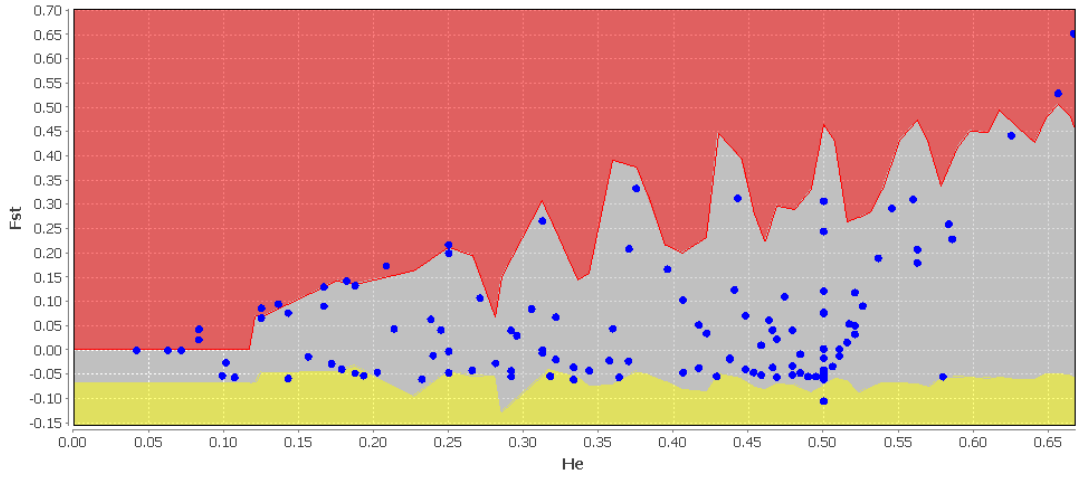
ISB vs. Hast outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 7 SNPs were detected as positive outliers (strong outliers).



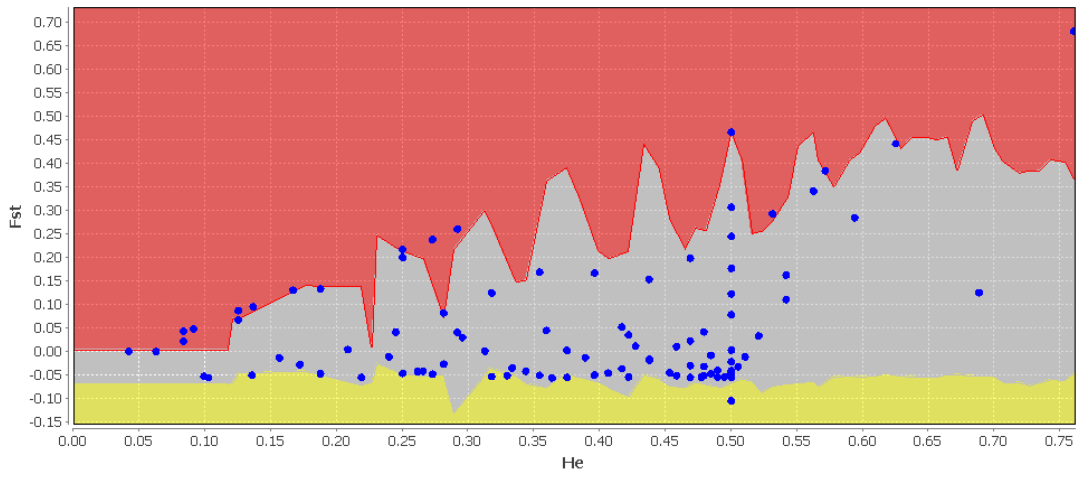
ISB vs. Har outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 9 SNPs were detected as positive outliers (strong outliers).



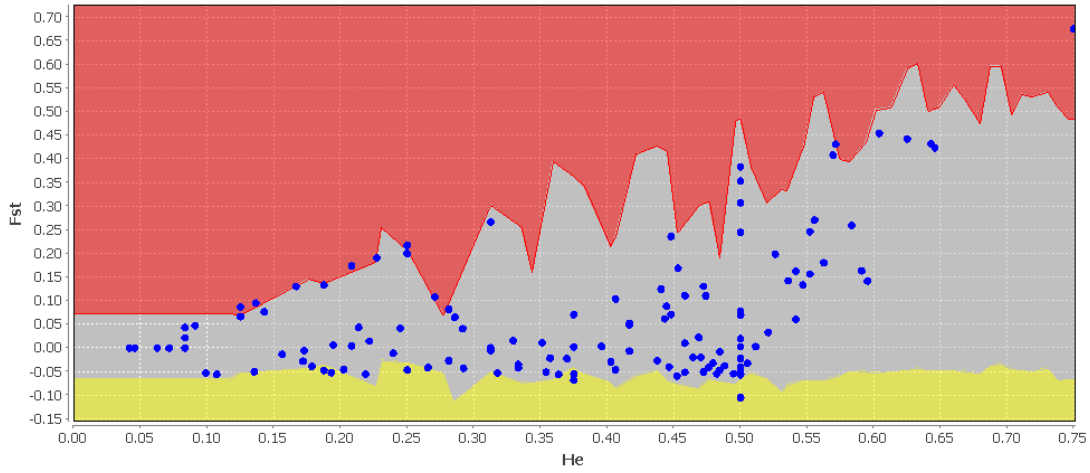
ISB vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 60 SNPs were detected as positive outliers (strong outliers).



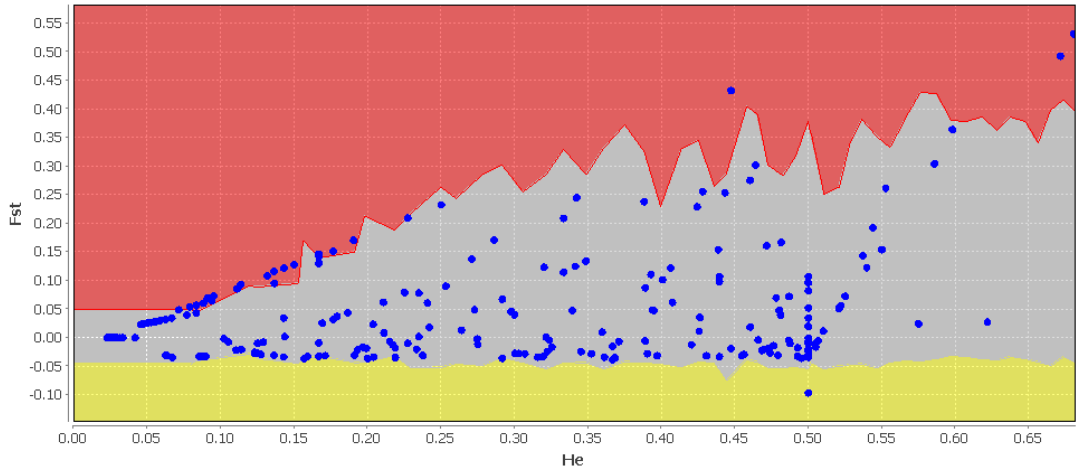
ISB vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 21 SNPs were detected as positive outliers, with 20 SNPs remaining significant outliers after FDR correction (strong outliers).



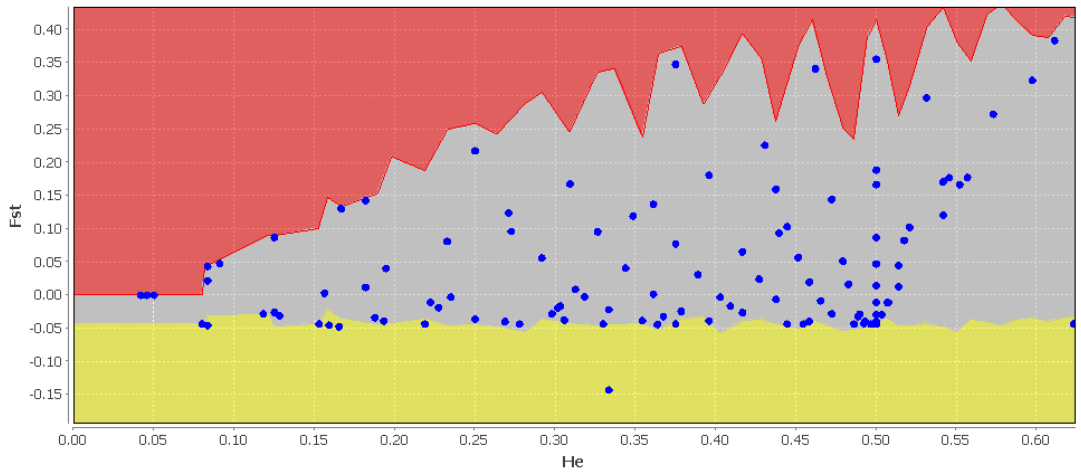
ISB vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 25 SNPs were detected as positive outliers (strong outliers).



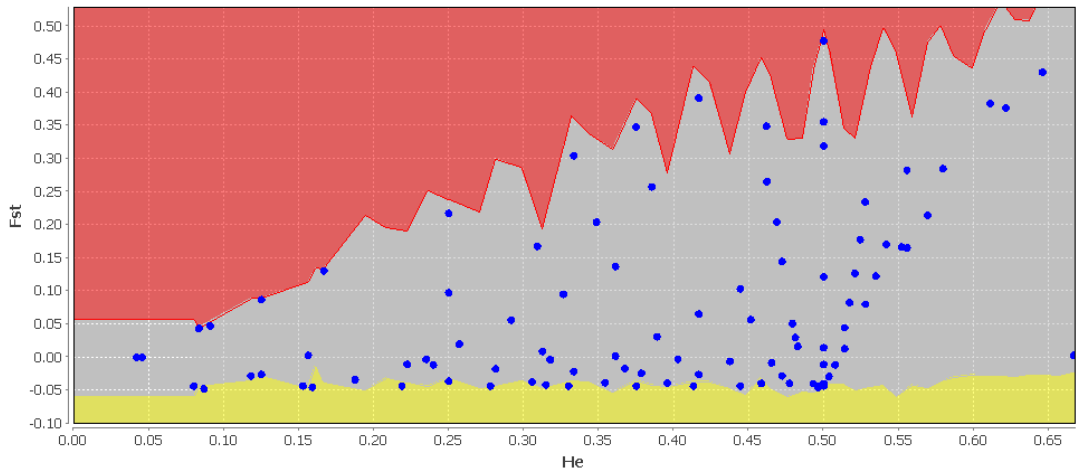
ISB vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 10 SNPs were detected as positive outliers, with 5 SNPs remaining significant outliers after FDR correction (strong outliers).



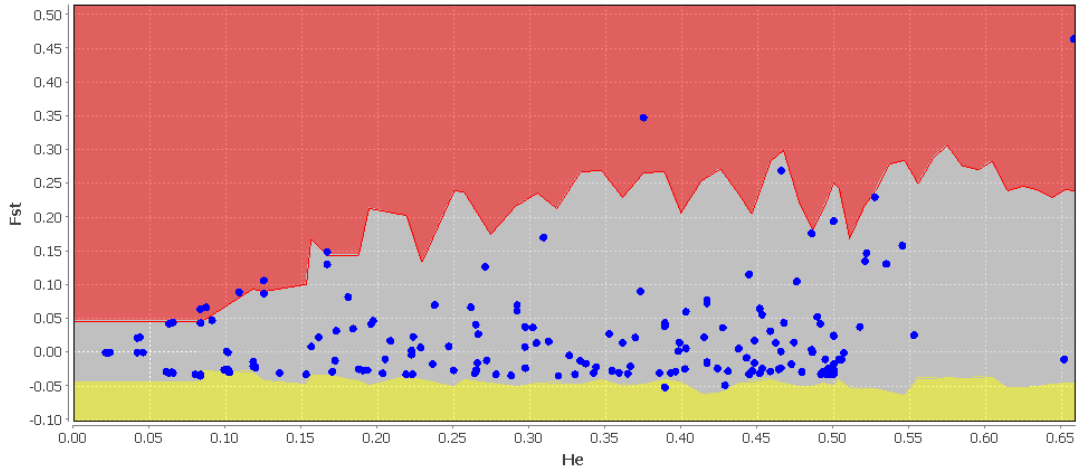
ISB vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 16 SNPs were detected as positive outliers (strong outliers).



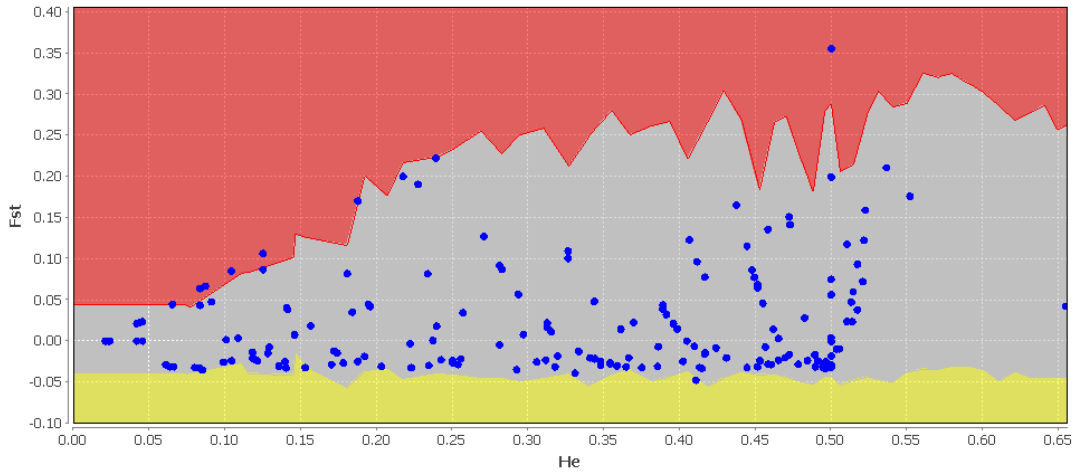
NQ vs. Brit outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). No SNPs were detected as positive outliers.



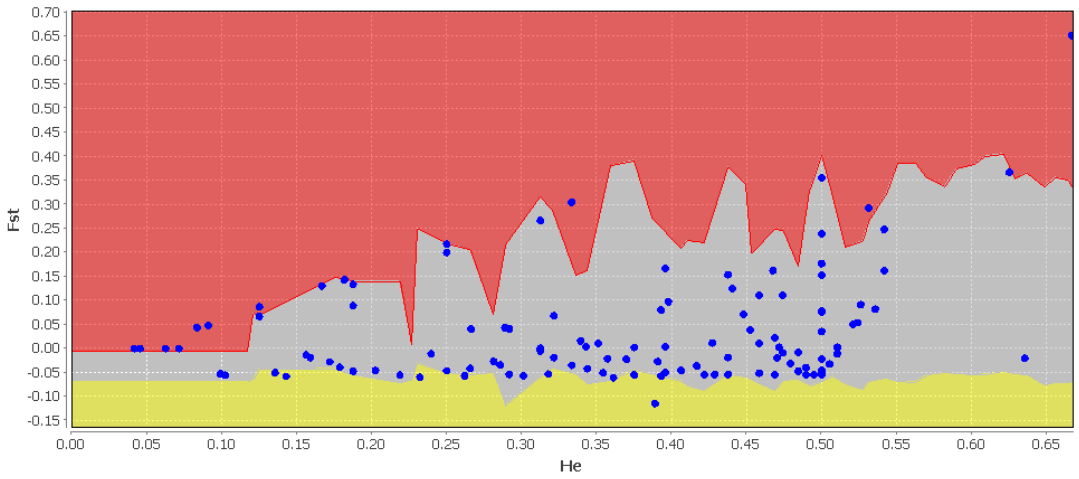
NQ vs. Hast outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). No SNPs were detected as positive outliers.



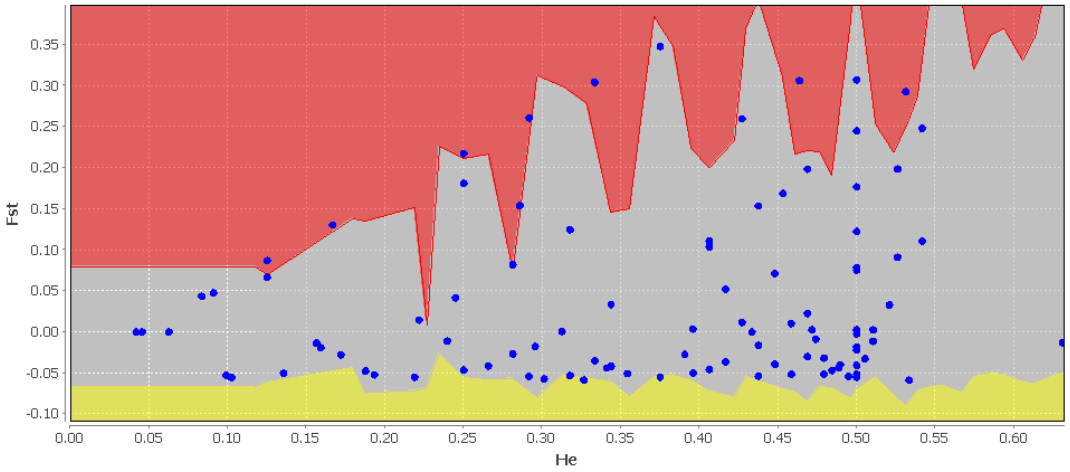
NQ vs. Har outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 24 SNPs were detected as positive outliers (strong outliers).



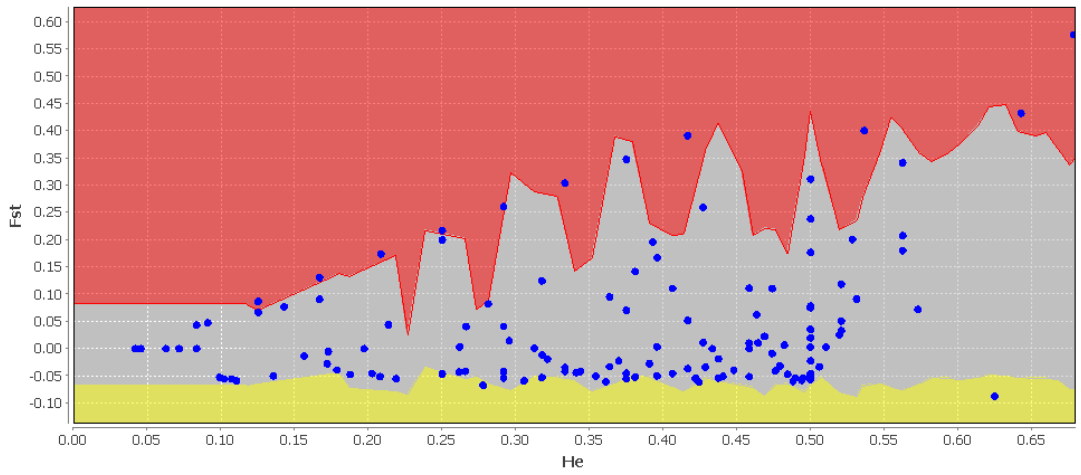
NQ vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 7 SNPs were detected as positive outliers (strong outliers).



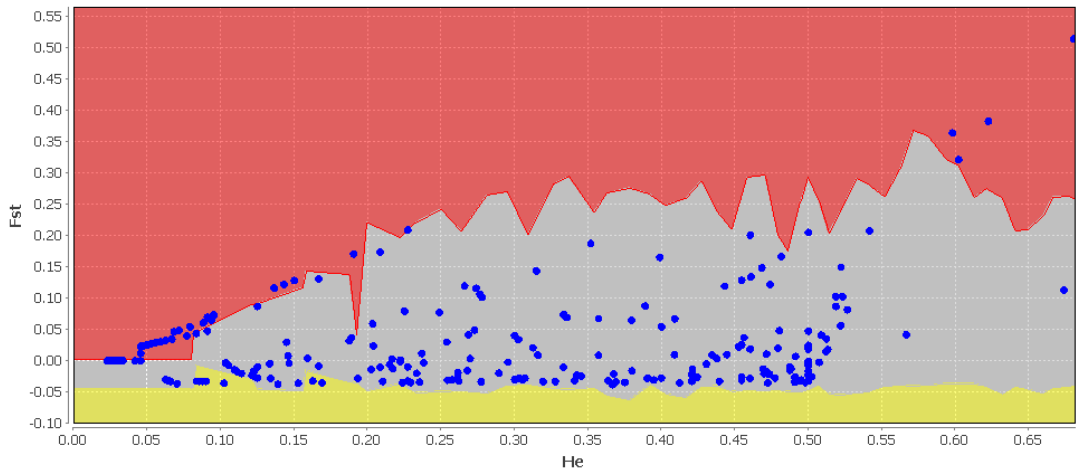
NQ vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 14 SNPs were detected as positive outliers (strong outliers).



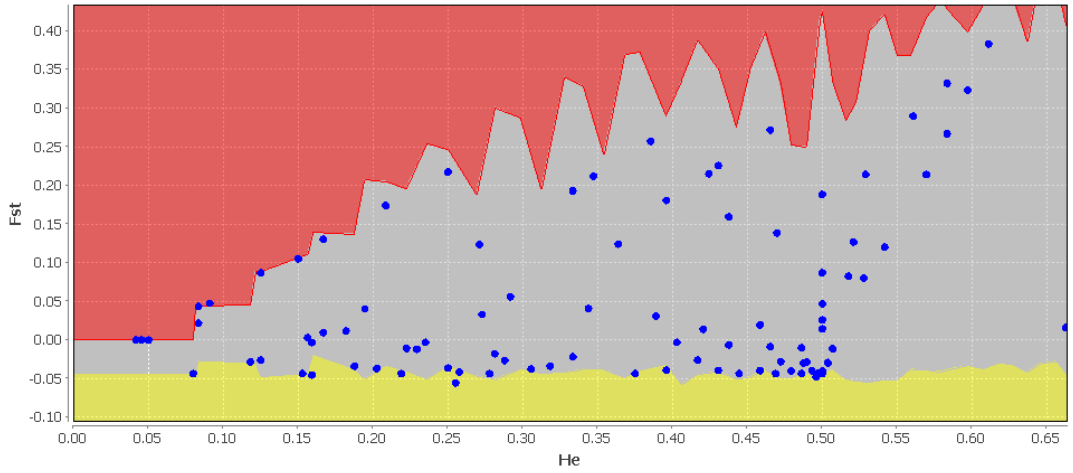
NQ vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 33 SNPs were detected as positive outliers (strong outliers).



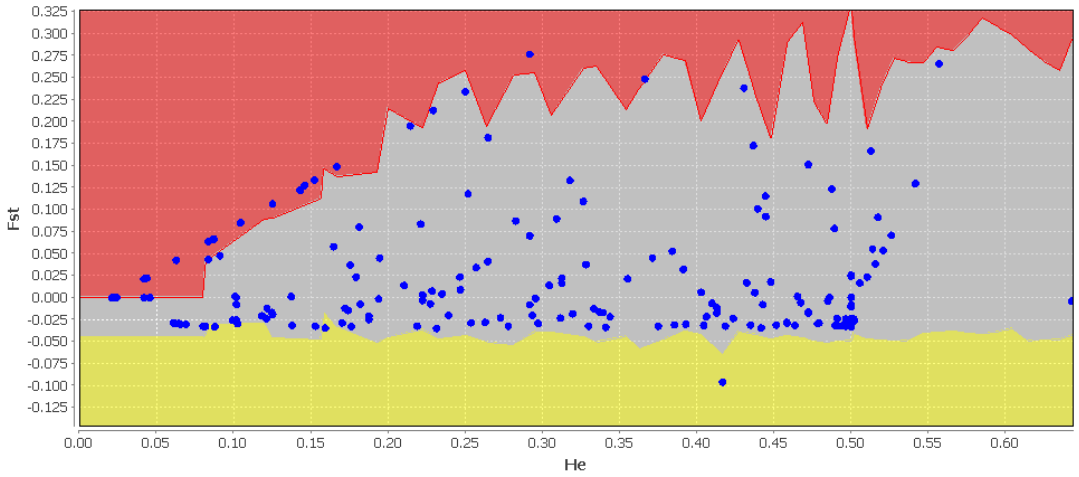
NQ vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 41 SNPs were detected as positive outliers (strong outliers).



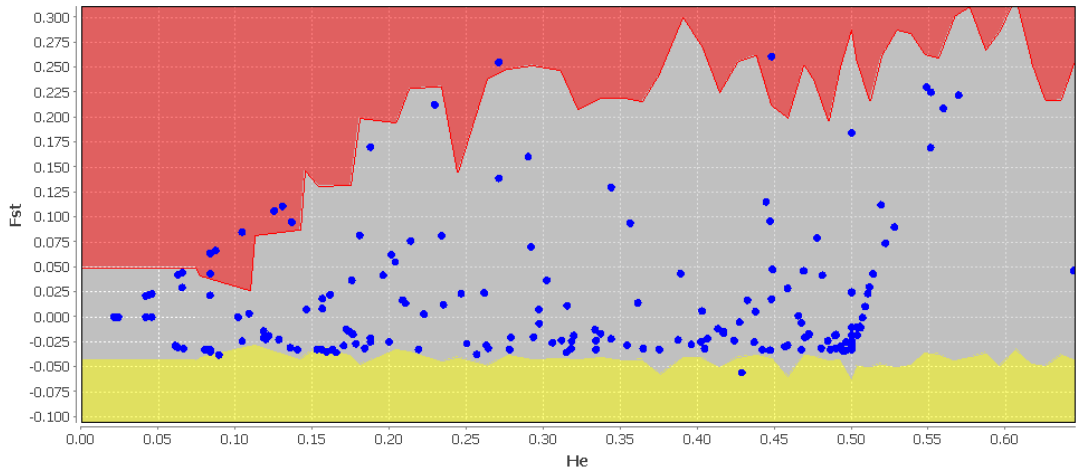
NQ vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 6 SNPs were detected as positive outliers (strong outliers).



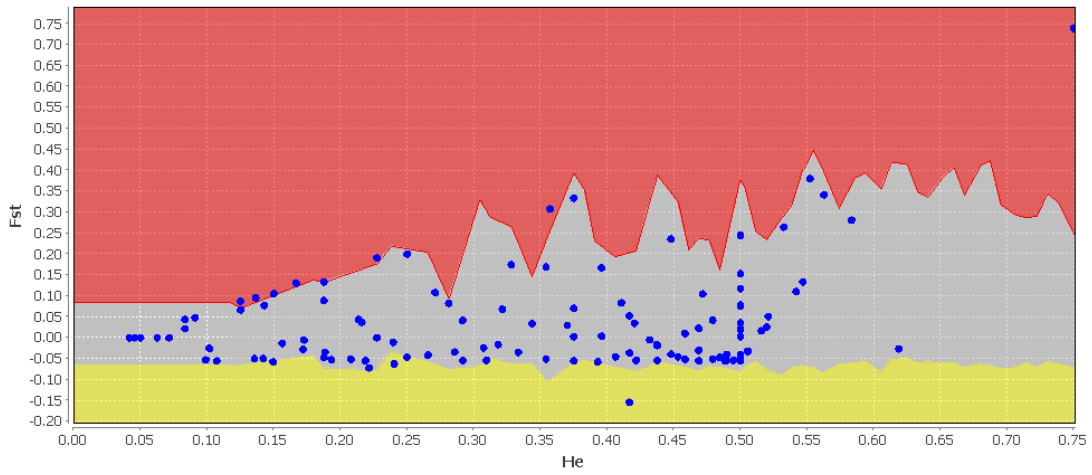
Brit vs. Hast outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 6 SNPs were detected as positive outliers (strong outliers).



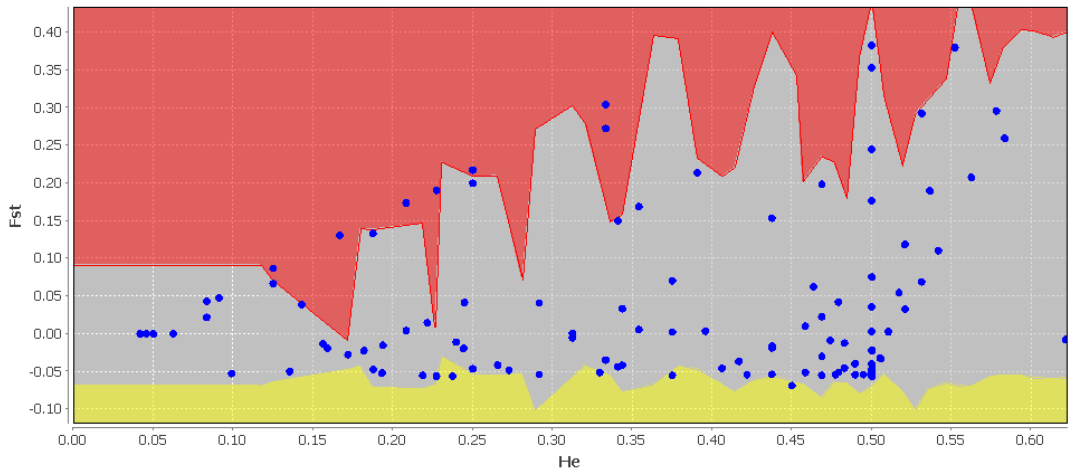
Brit vs. Har outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 8 SNPs were detected as positive outliers (strong outliers).



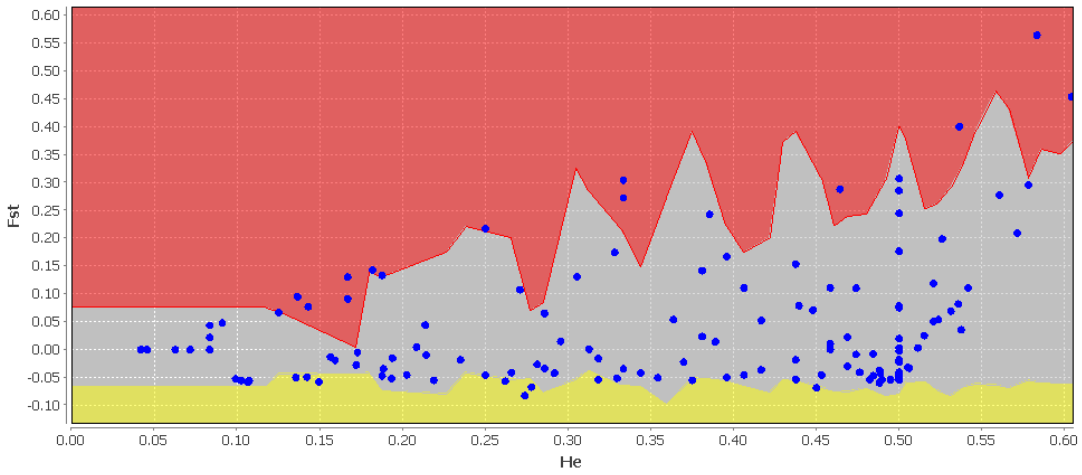
Brit vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 8 SNPs were detected as positive outliers (strong outliers).



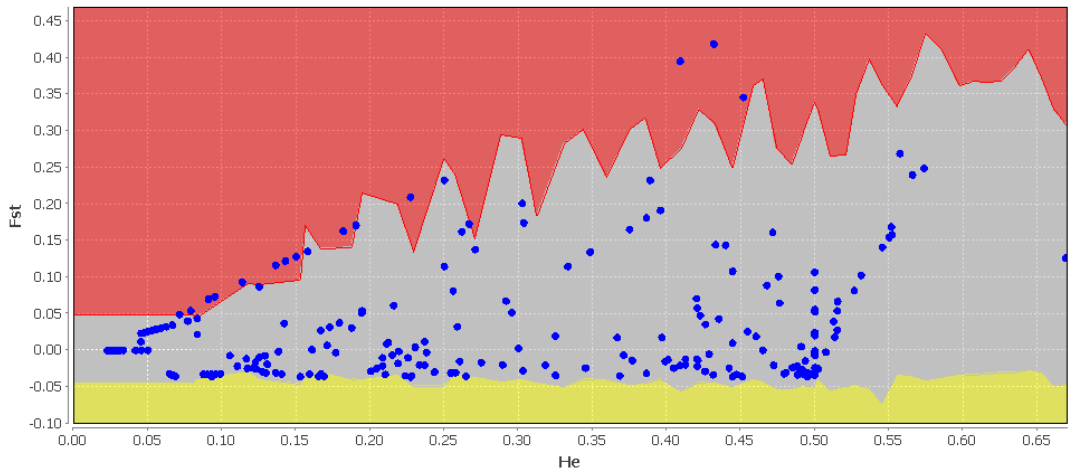
Brit vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 23 SNPs were detected as positive outliers (strong outliers).



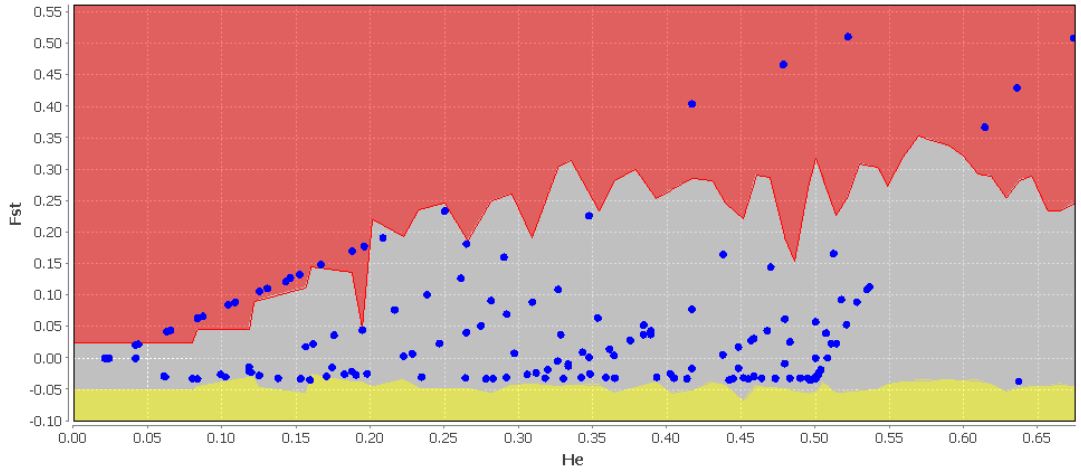
Brit vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 13 SNPs were detected as positive outliers (strong outliers).



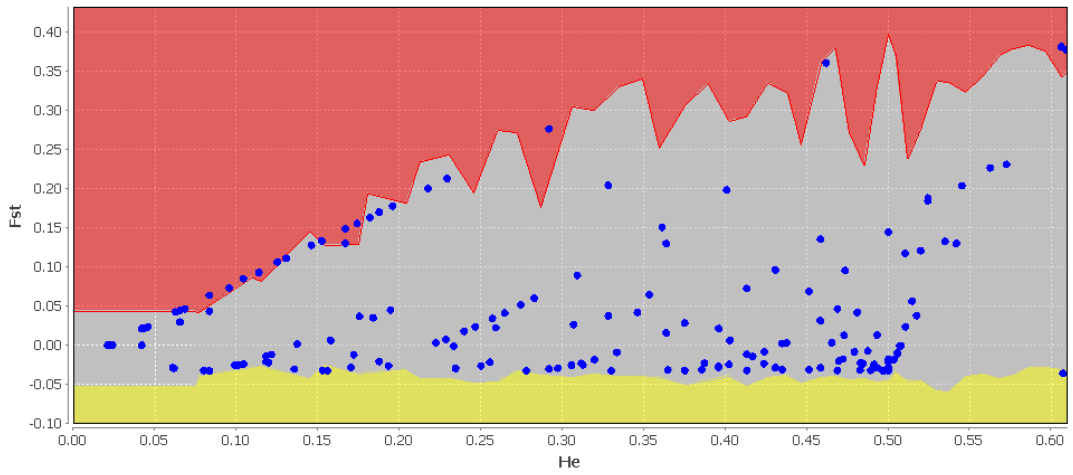
Brit vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 32 SNPs were detected as positive outliers (strong outliers).



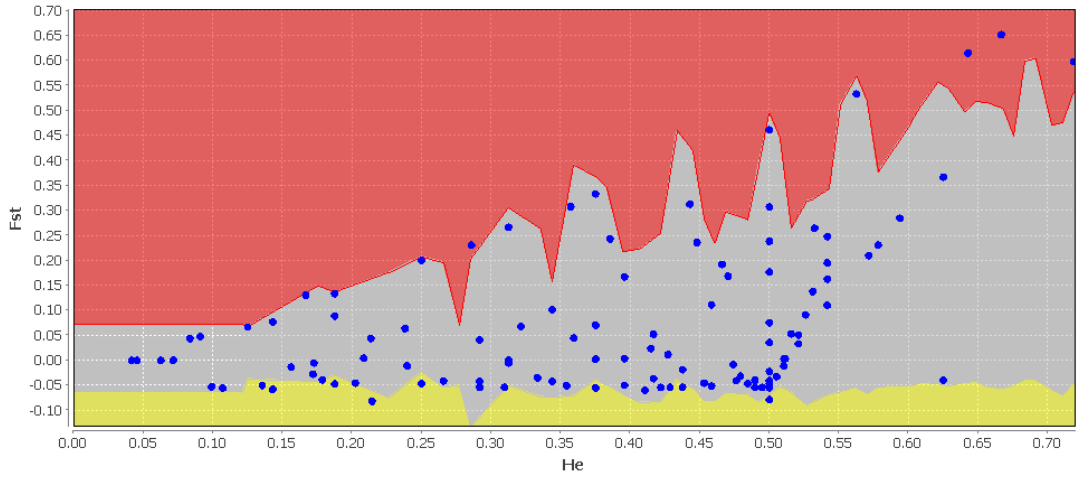
Brit vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 12 SNPs were detected as positive outliers (strong outliers).



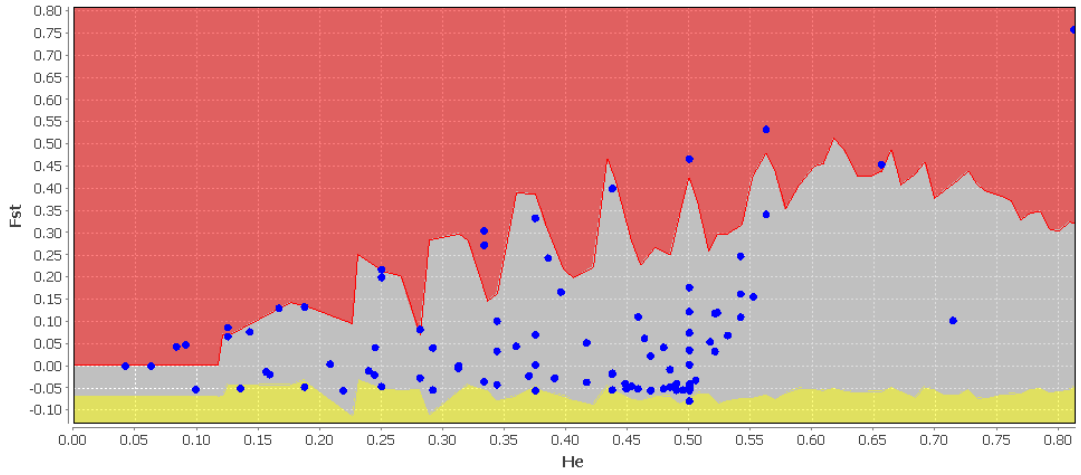
Hast vs. Har outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 13 SNPs were detected as positive outliers (strong outliers).



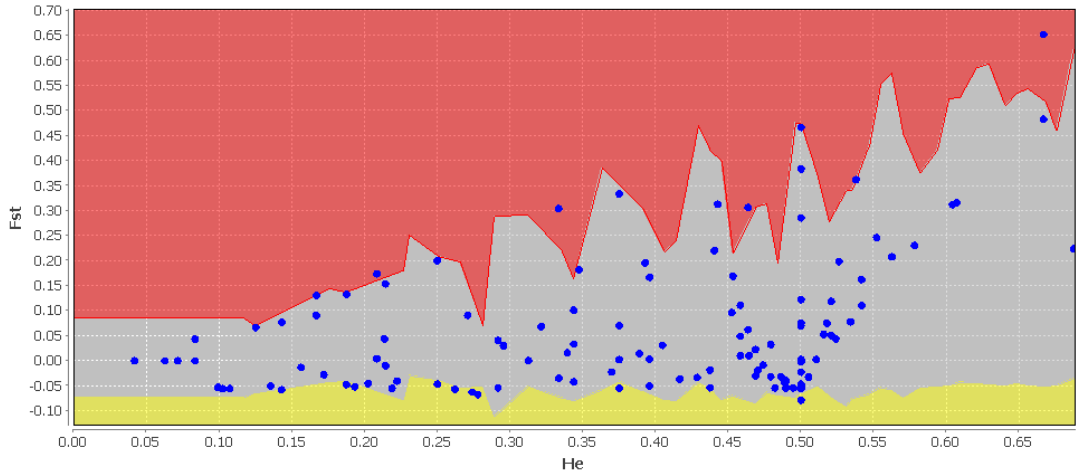
Hast vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 16 SNPs were detected as positive outliers, with 14 SNPs remaining significant outliers after FDR correction (strong outliers).



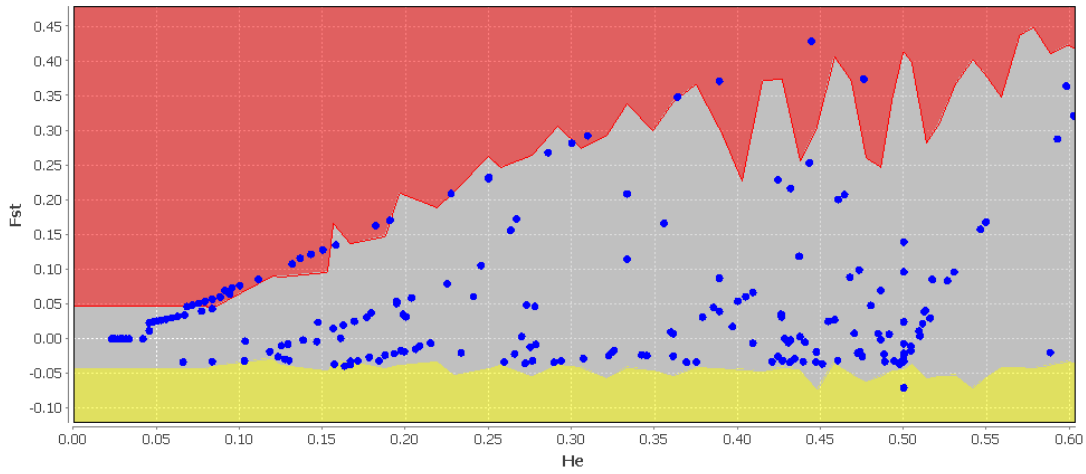
Hast vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 5 SNPs were detected as positive outliers (strong outliers).



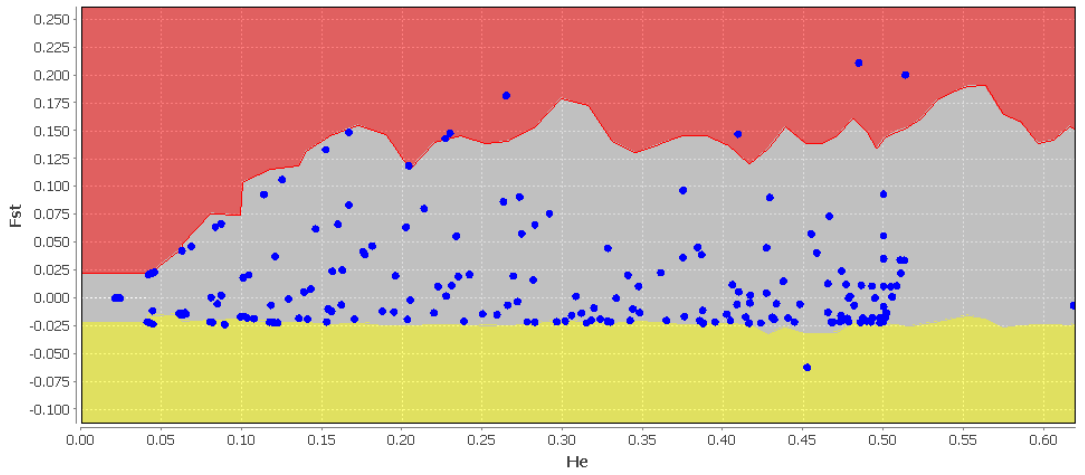
Hast vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 14 SNPs were detected as positive outliers, with 13 SNPs remaining significant outliers after FDR correction (strong outliers).



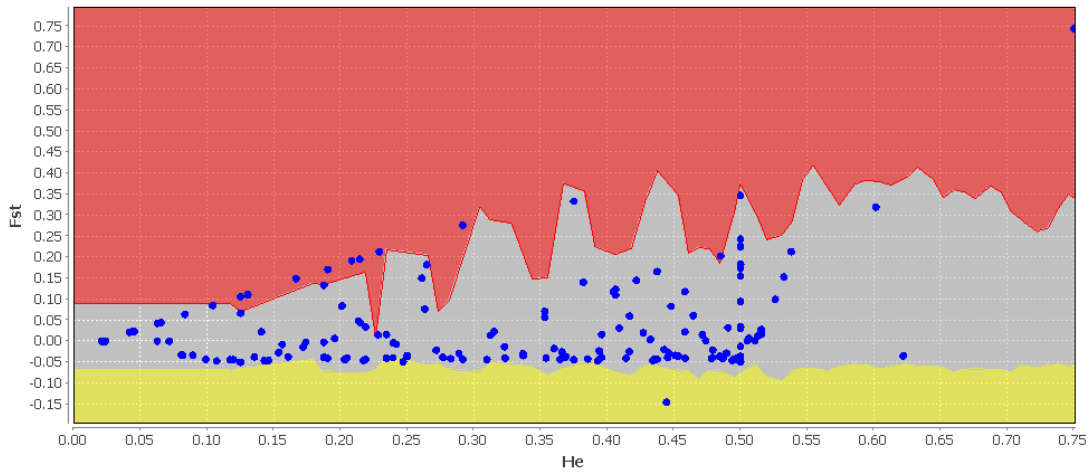
Hast vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 9 SNPs were detected as positive outliers, with 8 SNPs remaining significant outliers after FDR correction (strong outliers).



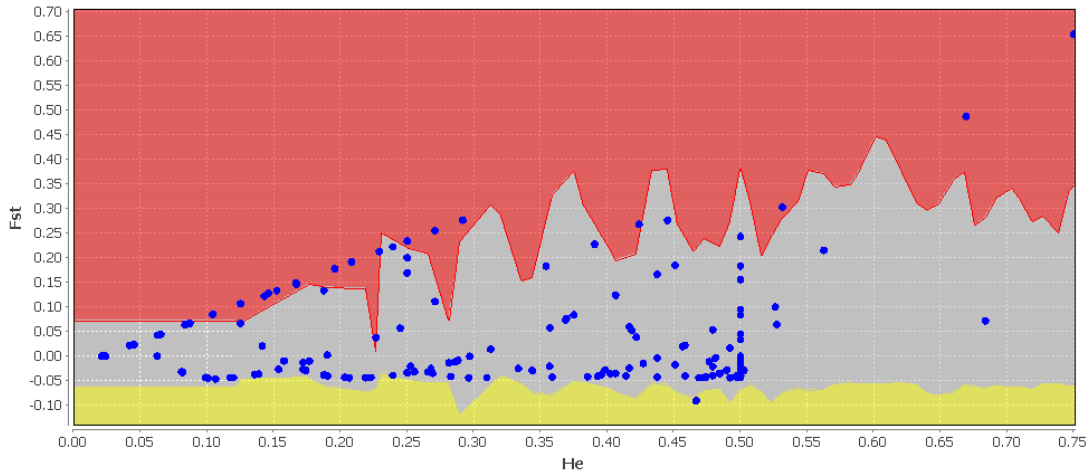
Hast vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 17 SNPs were detected as positive outliers (strong outliers).



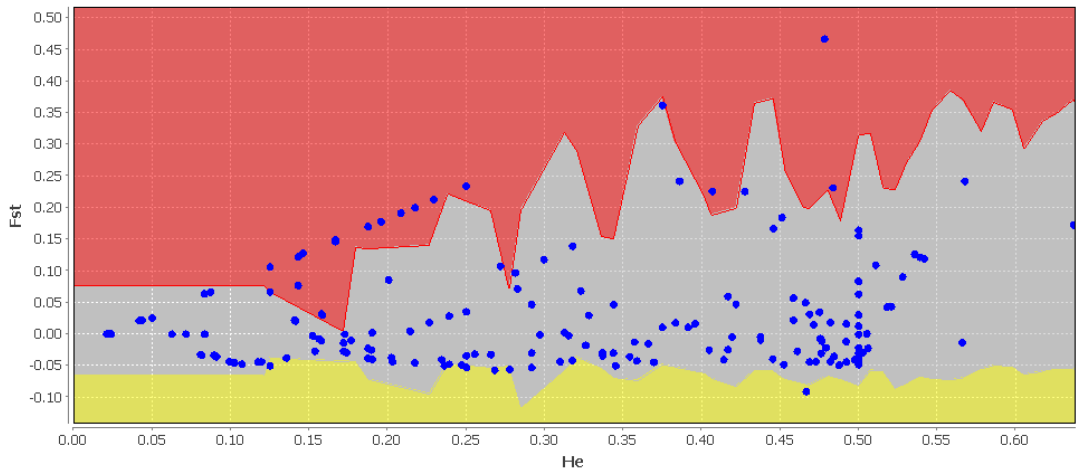
Har vs. Shet outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 4 SNPs were detected as positive outliers (strong outliers).



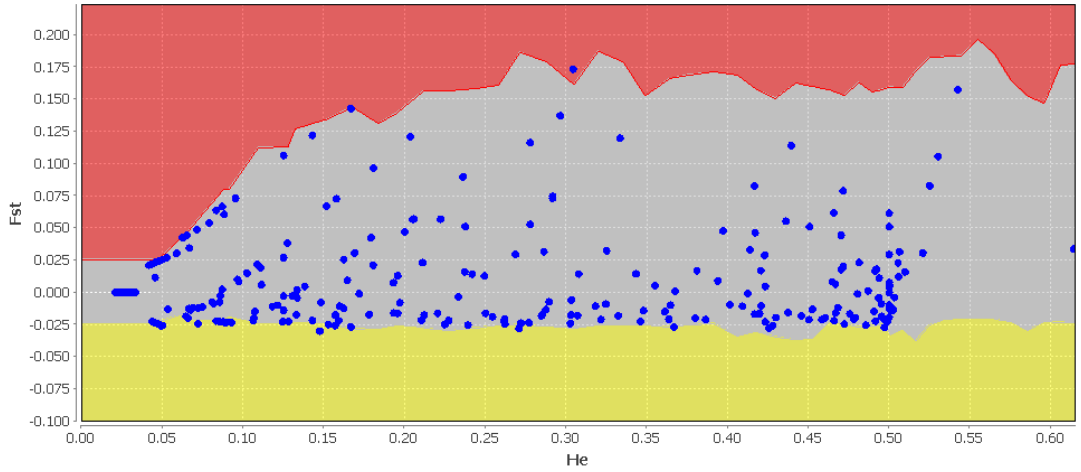
Har vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 16 SNPs were detected as positive outliers (strong outliers).



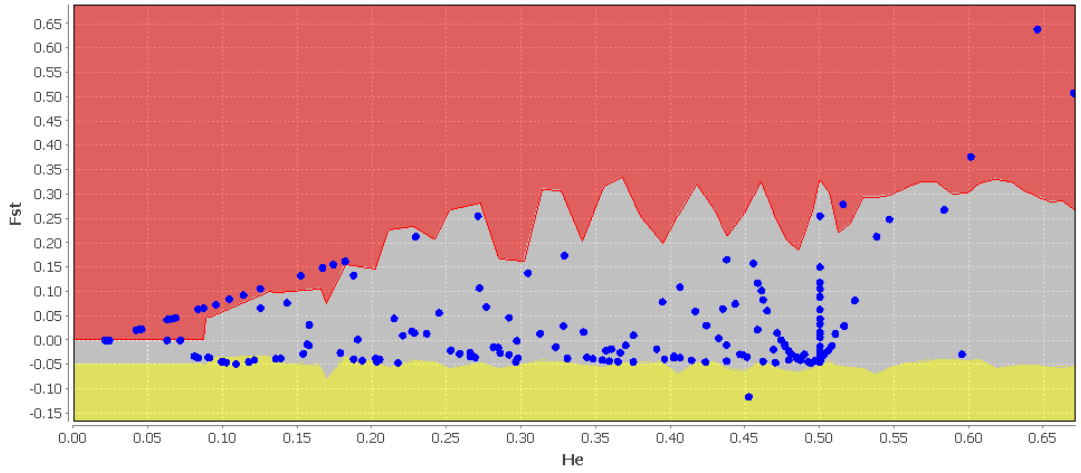
Har vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 26 SNPs were detected as positive outliers (strong outliers).



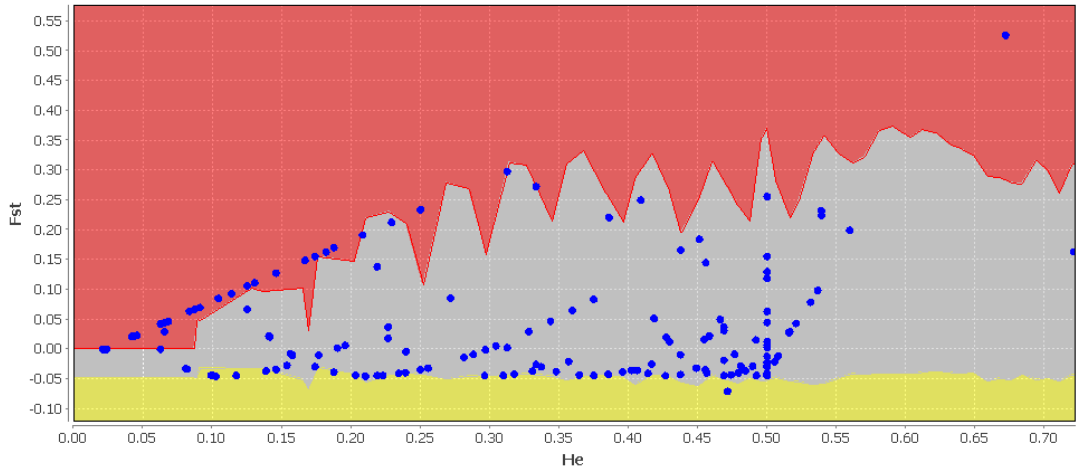
Har vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 12 SNPs were detected as positive outliers (strong outliers).



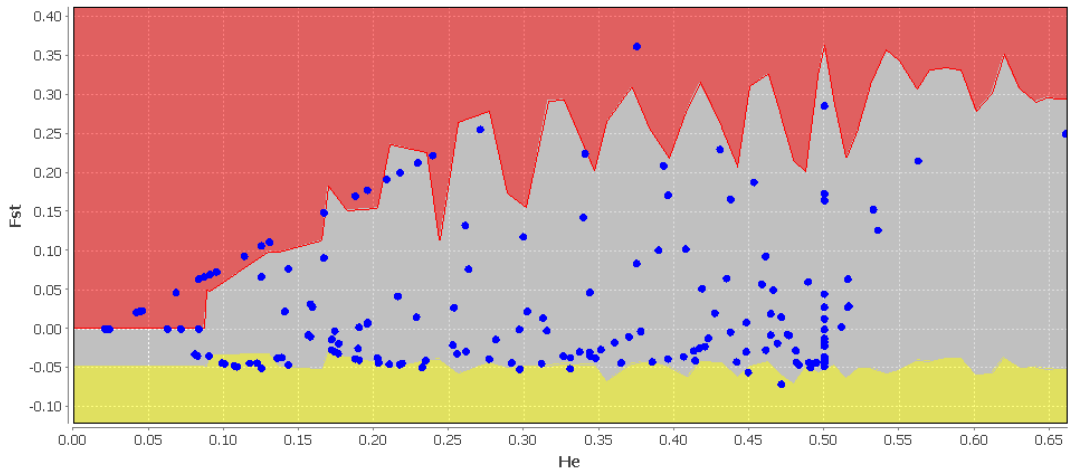
Har vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 6 SNPs were detected as positive outliers, with 5 SNPs remaining significant outliers after FDR correction (strong outliers).



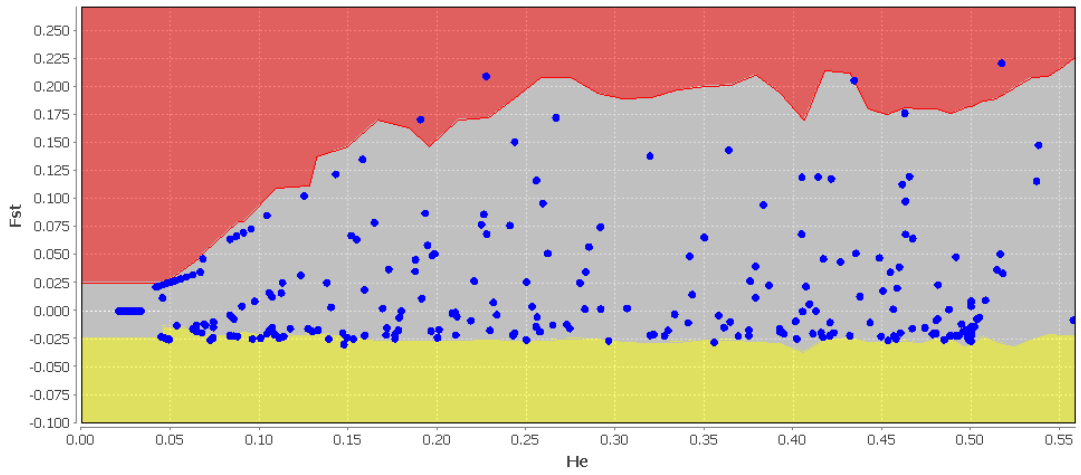
Shet vs. LL outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 24 SNPs were detected as positive outliers (strong outliers).



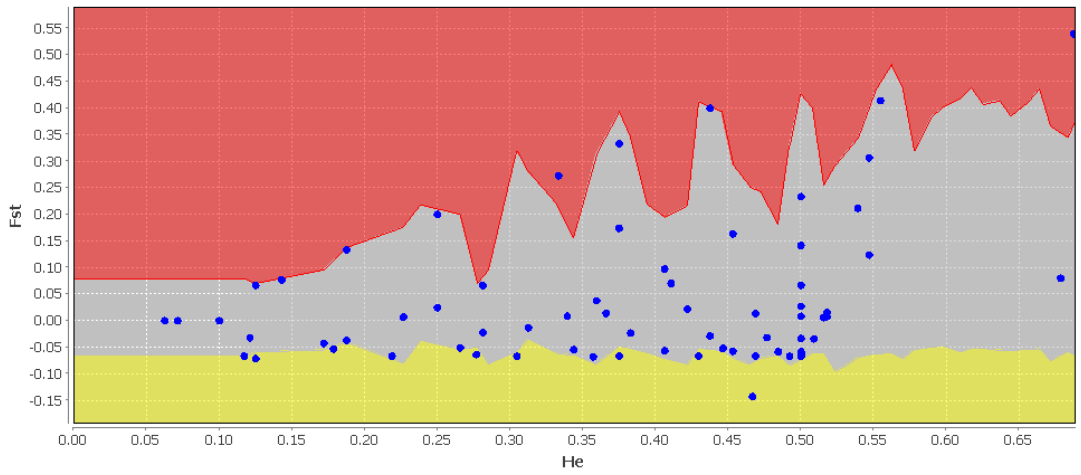
Shet vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 26 SNPs were detected as positive outliers (strong outliers).



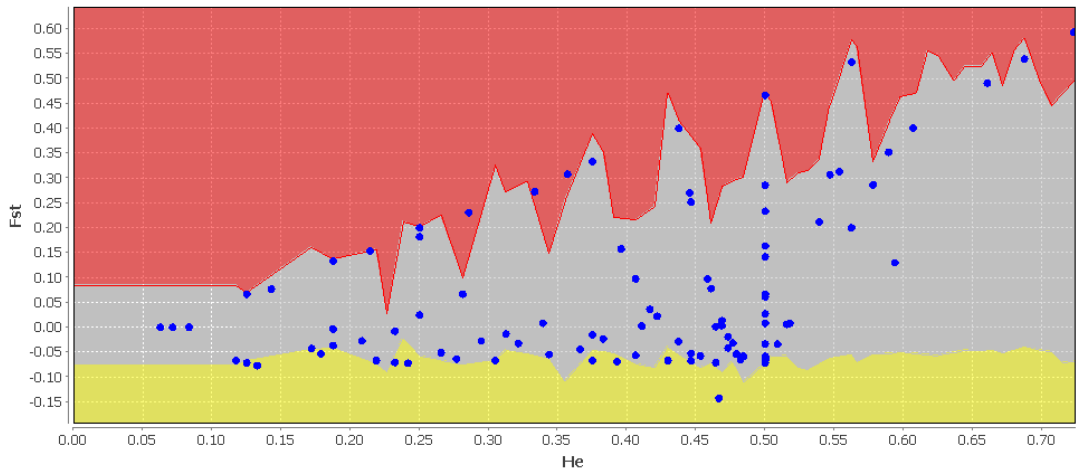
Shet vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 23 SNPs were detected as positive outliers (strong outliers).



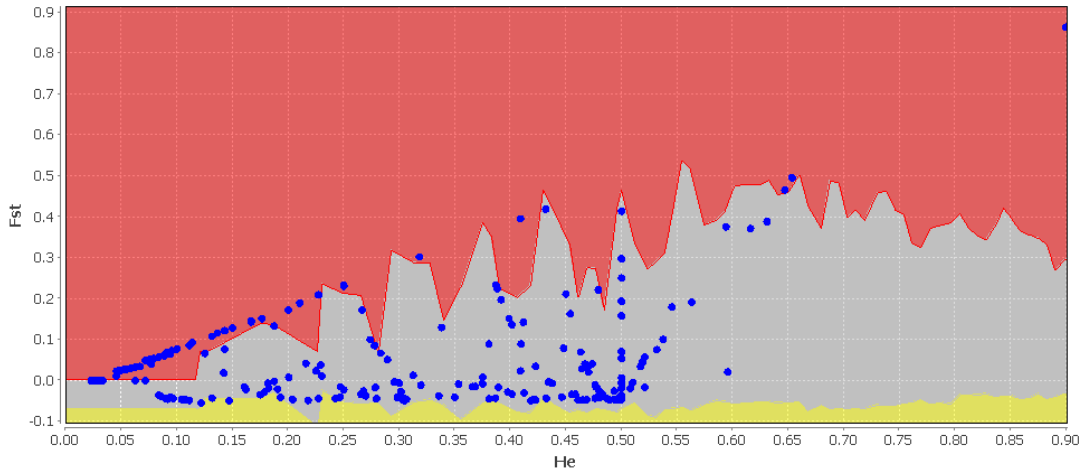
Shet vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 8 SNPs were detected as positive outliers, with 5 SNPs remaining significant outliers after FDR correction (strong outliers).



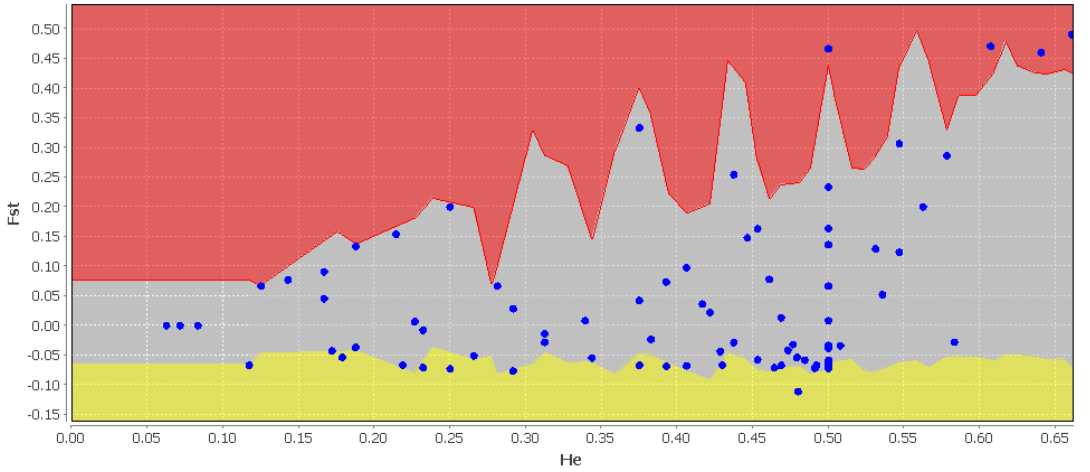
LL vs. LV outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 32 SNPs were detected as positive outliers (strong outliers).



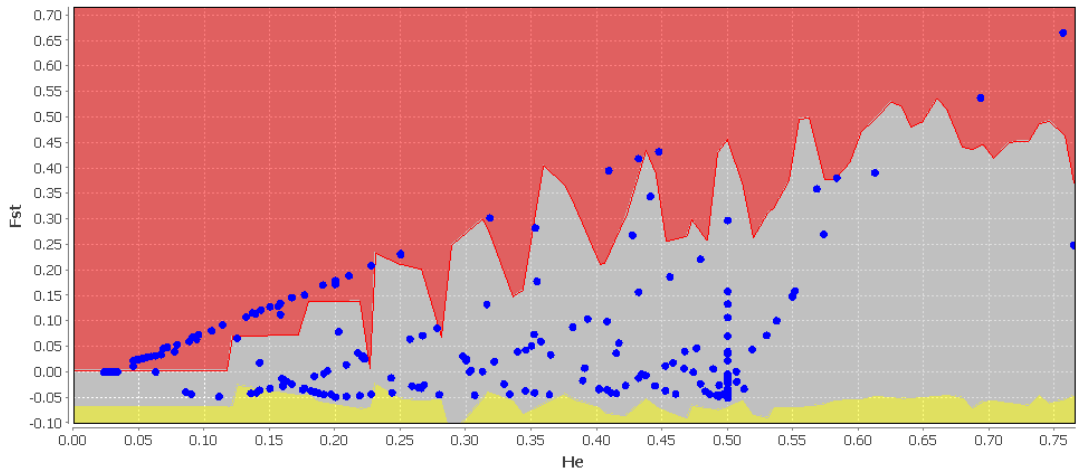
LL vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 5 SNPs were detected as positive outliers, with 4 SNPs remaining significant outliers after FDR correction (strong outliers).



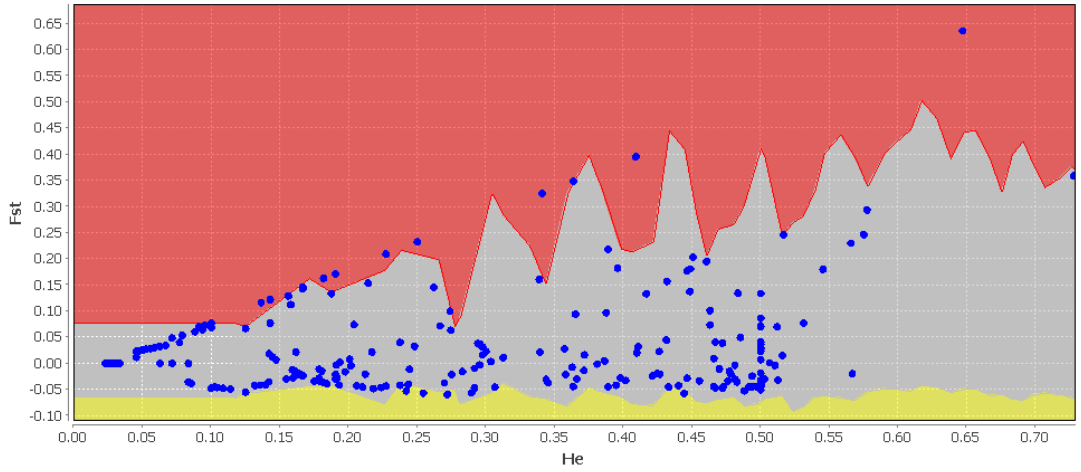
LL vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 37 SNPs were detected as positive outliers (strong outliers).



LV vs. G outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 56 SNPs were detected as positive outliers (strong outliers).



LV vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 43 SNPs were detected as positive outliers, with 42 SNPs remaining significant outliers after FDR correction (strong outliers).



G vs. Gul outlier test: 566 SNPs (blue dots) detected for *Cancer pagurus* in the NE Atlantic are considered to be candidates for balancing selection (yellow area), positive selection (red area) or neutral (grey area). 25 SNPs were detected as positive outliers (strong outliers).